U.S. Army Research Institute
for the Behavioral and Social Sciences

Research Report 1918

Sustainment of Individual and Collective Future Combat Skills: Modeling and Research Methods

Anna T. Cianciolo
Cognitive Performance Group

Brian T. Crabb, & Peter Schaefer
U.S. Army Research Institute

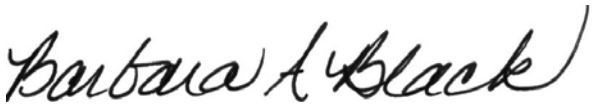
Steven Jackson, & Jeff Grover
Dynamics Research Corporation

January 2010

U.S. Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:



BARBARA A. BLACK, Ph.D.
Research Program Manager
Training and Leader Development Division

MICHELLE SAMS, Ph.D.
Director

Technical Review by

Scott Beal, U.S. Army Research Institute
Charles Johnson, Future Force Integration Directorate

NOTICES

DISTRIBUTION: Primary distribution of this Research Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-ZXM, 2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926.

FINAL DISPOSITION: This Research Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Research Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy)	2. REPORT TYPE	3. DATES COVERED (from. . . to)
January 2010	Final	July 2007– March 2009

4. TITLE AND SUBTITLE	5a. CONTRACT OR GRANT NUMBER
Sustainment of Individual and Collective Future Combat Skills: Modeling and Research Methods	W74V8H-04-D-0048/D.O. 0009
	5b. PROGRAM ELEMENT NUMBER 633007

6. AUTHOR(S)	5c. PROJECT NUMBER
Anna T. Cianciolo (Cognitive Performance Group); Brian T. Crabb, & Peter Schaefer (U.S. Army Research Institute); Steven Jackson, & Jeff Grover (Dynamics Research Corporation)	A792
	5d. TASK NUMBER 356
	5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Dynamics Research Corporation, 955 N. Wilson Rd., Suite E, Radcliff, KY 40160 Cognitive Performance Group, 3662 Avalon Park East Blvd, Suite 2023, Orlando, FL 32828 U.S. Army Research Institute for the Behavioral and Social Sciences, 121 Morande St., ATTN: DAPE-ARI-IK, Fort Knox, KY 40121-4141	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. MONITOR ACRONYM
U.S. Army Research Institute for the Behavioral and Social Sciences 2511 Jefferson Davis Highway Arlington, VA 22202-3926	ARI
	11. MONITOR REPORT NUMBER Research Report 1918

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Contracting Officer's Representative and Subject Matter POC: Brian T. Crabb

14. ABSTRACT (Maximum 200 words) Army commanders have insufficient time to train on every mission requirement and organizational standard. Mission essential task lists help to scope training requirements based on current performance. However, there presently is no way for unit trainers to systematically schedule their training based on *expected* performance. The ability to project training status outward, beyond current performance levels, would enhance decisions about scheduling training. The ARI has previously investigated skill retention in order to develop such a capability. Changes in the operational environment and in the theoretical understanding of human performance have created opportunities to advance ARI's research program and have necessitated that these advances be made to assist the warfighter. Our research assessed the implications of the contemporary operational environment for maintaining skilled performance in light of a host of theoretical factors thought to influence skill decay. We implemented our findings in a survey-based instrument to be used for rating individual and collective tasks on several of these retention factors. This paper describes the survey-based instrument, its development, and initial evaluation. In future work, task ratings assigned using this instrument will be compared to actual performance data in order to build and validate a quantitative model of individual and collective skill retention.

15. SUBJECT TERMS

Future Combat Systems Cognitive Themes

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	87	Ellen Kinzer Technical Publications Specialist 703-602-8049

Research Report 1918

Sustainment of Individual and Collective Future Combat Skills: Modeling and Research Methods

Anna T. Cianciolo
Cognitive Performance Group

Brian T. Crabb, & Peter Schaefer
U.S. Army Research Institute

Steven Jackson, & Jeff Grover
Dynamics Research Corporation

Fort Knox Research Unit
James W. Lussier, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926

January 2010

Army Project Number
633007A792

Personnel Performance and Training

ACKNOWLEDGEMENTS

We gratefully acknowledge the support and assistance of the following personnel and organizations who supported our effort: the Future Force Integration Directorate (FFID), Major General James Terry, Colonel Michael Wadsworth, Colonel Mathew Van Grutten (UK), Colonel Joshua Berisford, Lieutenant Colonel Ronald Daniel, Lieutenant Colonel Jason Rakocy, Garry Kendrik and Charlie Johnson; TRADOC Analysis Center Whitesands Missile Range, Eric Johnson, Kevin Wainer, Sara Krondak and Greg Kent; the 2nd Combined Arms Battalion, Lieutenant Colonel Edward House and Major Randy Alfredo; the Close Combat Tactical Trainer, Rocky Hanlin and Dave Hieatt and Lieutenant Colonel (Ret.) Roy Krueger (Program Manager), Dynamics Research Corporation.

SUSTAINMENT OF INDIVIDUAL AND COLLECTIVE FUTURE COMBAT SKILLS: MODELING AND RESEARCH METHODS

EXECUTIVE SUMMARY

Research Requirement:

Army commanders have insufficient time to train individuals and collectives on every mission requirement and organizational standard. Mission essential task lists help to scope training requirements based on current individual and collective performance, however there presently is no way for unit trainers to systematically schedule their training based on *expected* performance. The ability to project training status outward, beyond current training status, would inform decisions about scheduling pre-deployment and refresher training. The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) has previously investigated skill retention to develop such a capability. In the two decades following this research program, however, there have been numerous changes both in the operational environment and in the theoretical understanding of skill retention. All of these changes simultaneously provide opportunities to advance ARI's skill retention research program and necessitate that these advances be made to assist the warfighter. The primary purpose of our research program was to revise ARI's existing research product, the User's Decision Aid (UDA) survey, and then apply it to modeling the sustainment of individual and collective Future Combat Systems (FCS) Spin Out (SO) skills. A secondary purpose of our research was to explore measures that would enable greater sensitivity in collective performance assessment. A third purpose of our research was to identify future research issues related to the training, performance, and retention of FCS skills.

Procedure:

The present research program emerged from a partnership between ARI and the Future Forces Integration Directorate (FFID), who shared an interest in understanding the unit training and performance implications associated with FCS technologies. The FFID permitted on-site data collection and participated actively in facilitating access to personnel, facilities, and data. The scope of our effort necessitated that we conduct the lion's share of our investigations by integrating our efforts with those of FFID, working with the tasks they were training, sampling from the performance data they collected, and using their observations about how exercise events unfolded.

To study FCS skill sustainment, we revised the UDA survey to address known limitations, to enhance it in response to recent reviews of skill retention research and changes in the operational environment, and to make it applicable to FCS individual and collective tasks. We selected a set of FCS individual and collective tasks to model, asking subject matter experts to rate the tasks using our revised UDA. We collected performance data on these tasks during the live exercise segment of FFID's SO integration mission.

To study FCS collective skill assessment, we researched and developed a measure of FCS-enabled troop-leading procedures (TLP) called the TLP Observer Checklist. Our intent was

not to create a technologically sophisticated assessment tool but to identify the competencies that should be measured, to specify a feasible observer-based assessment procedure, and to explore the human factors that influence the adoption of refined performance measures. To go a level deeper than TLP doctrinal task steps/performance measures, we extended existing observer-based methods for assessing tactical cognitive expertise (Phillips, Ross, & Shadrick, 2006). We sought to retain the many state-of-the-art characteristics of existing methods while simultaneously extending them to assess collective performance demonstrated during planning or collaborative problem solving. We applied our TLP Observer Checklist to observing a small set of simulation-based planning exercises conducted as part of the train up to FFID's live FCS integration exercises.

Findings:

The context we operated in, necessitated by a wartime environment, was characterized by a wealth of information about SO tasks, but also by competing requirements for performance data (behavioral research vs. technology integration). Contrasting with the previous ARI studies of skill retention, the SO tasks that we examined were not yet fully established and documented in doctrinal training manuals. Consequently, the nature of many of these tasks changed, sometimes substantively, as part of the natural course of the technology integration carried out by FFID. We also conducted performance data collection in a somewhat more uncontrolled context, relying on the assessment data provided by observer/trainers whose goal was to ensure that all units achieved proficiency on the FCS tasks evaluated. Our revised versions of the UDA demonstrated very high inter-rater reliability. However, further analyses of skill sustainment were not possible using the available performance data, which lacked quantity and variability. Similarly, it was determined that the TLP Observer Checklist format was easy to use, although the training performance data available were insufficient to analyze the psychometric properties of the measure. Lessons learned in our research pointed the way to conducting more productive quantitative analysis in the context of ongoing technology integration missions. These lessons learned served as the basis for our research plans to be conducted in a future phase of our investigations.

Utilization and Dissemination of Findings:

Even in its interim state, the present research program represents a significant advancement in the understanding, assessment, and exploration of collective skill retention. It builds on recognized best practices in skill retention research, begun by ARI in the mid-1980's, and addresses specific areas where research and development is especially needed. With refined models of collective skill retention, this research program will be the first of its kind in producing useful quantitative analysis of the factors that influence collective skill retention. In addition, through enhanced performance assessment techniques, this work will provide useful methods for conducting psychological research in the context of ongoing training and evaluation exercises necessary to maintain Army readiness.

SUSTAINMENT OF INDIVIDUAL AND COLLECTIVE FUTURE COMBAT SKILLS: MODELING AND RESEARCH METHODS

CONTENTS

Page

Introduction....1
Research Purpose and Goals....3
Overview of this Report....3

Overview of the Research Program....3
Research Context....4

Skill Retention Research....5
Background....5
The Present Research....6
Revising the User's Decision Aid....6
The Trainer's Decision Aid....11
Spin Out Tasks Selected....11
Procedure....14
Results....14
Lessons Learned....16

Collective Skill Measurement Research....17
Design of the TLP Observer Checklist....18
TLP Observer Checklist Development & Administration Process....21
TLP Observer Checklist Lessons Learned....21

Next Steps....23
Skill Retention Research....23
Collective Skill Measurement Research....25

Conclusions....26

References....27

Appendix A. UDA Example Item....A-1
Appendix B. TDA Survey – Individual Tasks....B-1
Appendix C. TDA Survey – Collective Tasks....C-1
Appendix D. Individual and Collective Tasks Selected....D-1
Appendix E. TLP Observer Checklist....E-1

Page

List of Tables

Table 1. Comparison of Items on the User's Decision Aid (UDA) and the Trainer's Decision Aid (TDA) 12
Table 2. Inter-rater Reliabilities on Selected TDA Questions 16
Table3. Troop Leading Procedure Task Steps for Different Platoons Equipped with Spin Out Technologies....19

List of Figures

Figure 1. Simplistic representation of projected task performance based on initial training status 1
Figure 2. Simplistic representation of projected task performance based on known properties of skill retention....2
Figure 3. Progression of training, test, and evaluation exercises conducted by the 5/1 AD (AETF)....4
Figure 4. Schematic depiction of actual and predicted performance trends (Rose, Czarnolewski et al., (1985)....7
Figure 5. Conceptual latent variable model of skill retention....25

SUSTAINMENT OF INDIVIDUAL AND COLLECTIVE FUTURE COMBAT SKILLS: MODELING AND RESEARCH METHODS

Introduction

It is widely recognized within the Army that there is insufficient time to train individuals and collectives (e.g., a team, crew, squad, platoon, company, etc.) on every mission requirement and organizational standard (e.g., Wong, 2002). Leaders at all echelons must make tough decisions as to what directed training they will conduct, and much of a Soldier's knowledge and skill is acquired on the job. Mission essential task lists help to scope training requirements based on current individual and collective performance. However, there presently is no way for unit trainers to systematically schedule their training based on *expected* performance.

Consider a unit trainer with four mission essential tasks. Figure 1 below is a simplistic depiction of the underlying assumption he or she must use to prioritize pre-deployment training on these tasks: once a task is trained to proficiency, that proficiency is retained through deployment and mission execution. Under this assumption, ensuring that all mission essential tasks are trained to proficiency in garrison is necessary and sufficient for sustaining operational readiness throughout deployment.

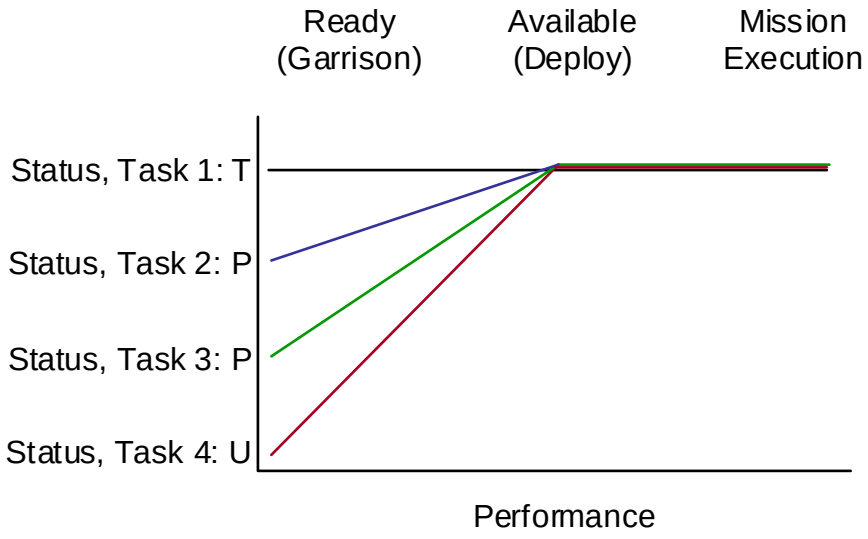


Figure 1. Simplistic representation of projected task performance based on initial training status. Note. T = Trained; P = Partially Trained; U = Untrained

Figure 2 depicts an alternative assumption on which to base pre-deployment training priorities: even though a task is trained to proficiency, skill is subject to decay in the absence of opportunities to perform, particularly within the first two months of no practice. The lines in this figure account for varying degrees of skill retention (depending on the task) and reflect the fact that even if a task is trained to proficiency in garrison, training status may decrease prior to deployment due to skill decay. An important implication of this figure is that the ability to

project training status outward, beyond current training status, would inform decisions about scheduling pre-deployment training.

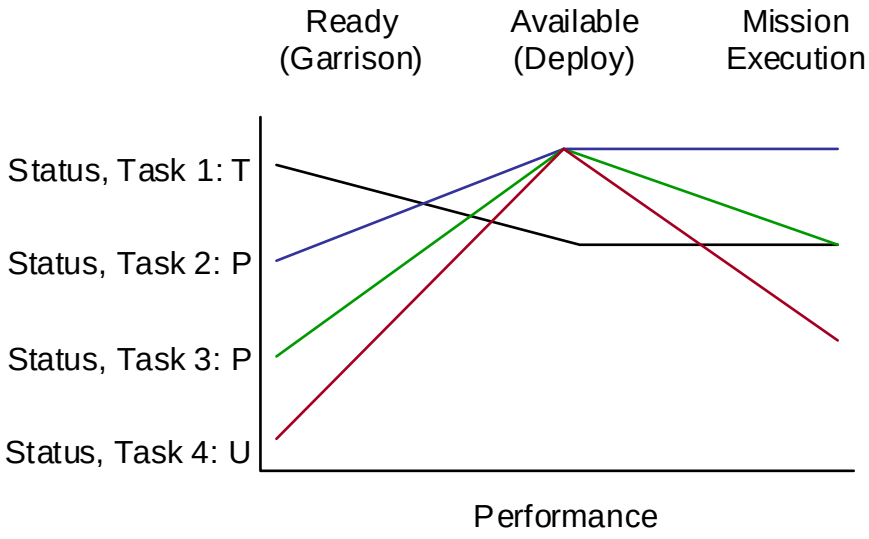


Figure 2. Simplistic representation of projected task performance based on known properties of skill retention.

This is not to say that unit trainers are unaware that skills decay over time, or that Soldiers do not actively conduct refresher training to maintain readiness levels. Rather, the point is that there is no systematic basis for making judgments about the rate of skill decay for mission essential tasks. These judgments are necessary for optimally scheduling refresher training requirements throughout the Army Force Generation cycle.

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) has previously investigated skill retention to develop a method for making such judgments (Macpherson, Patterson, & Mirabella, 1989; Rose, Czarnolewski, Gragg, Austin, & Ford, 1985; Rose, Radtke, Shettel, & Hagman, 1985a, 1985b; Sabol, Chapell, & Meiers, 1990; Wisher, Sabol, Sukenik, & Kern, 1991). A product of this research was the User's Decision Aid (UDA) (Rose, Radke et al., 1985a, 1985b), which was a survey-based method by which unit trainers could rate the characteristics of individual tasks and use this information to determine projected retention curves (i.e., the percentage of individuals in the unit who achieve "GO" status on the task at several points in time following no practice). The UDA was structured according to a theory-based model of individual skill retention, developed and refined over years of intensive data collection and analysis. Finished retention curves for several tasks (mobile subscriber equipment operator skills) were provided in one of the ARI reports documenting this research program (Sabol et al., 1990).

Over the past two decades, there have been numerous changes both in the operational environment and in the theoretical understanding of skill retention. For instance, gradual implementation of Future Combat Systems (FCS) technologies has introduced (and will continue to introduce) new mission tasks and new performance requirements in the areas of collaborative

information display usage and human-machine system integration. These tasks were not modeled in the previous ARI research. In addition, new theory in the areas of individual and collective skill retention has shed light on additional factors to consider when making predictions about skill decay. All of these changes simultaneously provide opportunities to advance ARI's skill retention research program and necessitate that these advances be made to assist the warfighter.

Research Purpose and Goals

The primary purpose of the present research was to revise and then apply the UDA to model the sustainment of individual and collective FCS Spin Out (SO) 1 skills. The UDA was to be updated by incorporating recent advancements in the theoretical understanding of individual skill retention and findings from initial explorations of collective skill retention. The revised UDA was then to be validated with individuals and collectives conducting FCS SO tasks. The goal of this effort was to produce performance retention predictions for the FCS skills studied as well as a general framework for predicting skill decay on related FCS tasks.

A secondary purpose of the present research was to explore measures that would enable greater sensitivity in collective performance assessment. Such measures could be used to refine retention predictions as well as provide more diagnostic information on the outcomes of collective performance due to training or other interventions.

Overview of this Report

This interim research report summarizes the current status of our research, details key lessons learned, and describes the steps necessary to complete the program. First, a general overview of our research program is provided, followed by detailed descriptions of both the skill sustainment and collective skill measurement investigations. The detailed descriptions include the investigative method used, the research procedure, materials, and products, data analyses and results, and lessons learned. Next, our planned future tasks are presented and their link to the lessons learned is specified. We also summarize the research issues we identified as part of our work. This report concludes with a brief statement on the impact of the current work and the potential utility of the planned future research.

Overview of the Research Program

The present research program emerged from a partnership between ARI and the Future Forces Integration Directorate (FFID), a relatively new component of the Army Capabilities Integration Center located at Fort Bliss, TX. The ARI and FFID shared an interest in understanding the unit training and performance implications associated with FCS technologies. Data collection at Fort Bliss with respect to advancing knowledge about FCS skill sustainment was permitted by FFID, and FFID participated actively in facilitating access to personnel, facilities, and data. We sought to integrate our efforts as closely as possible with FFID's ongoing activities such that our research presence was transparent to FFID personnel and affiliates. Obtaining this objective involved an on-site project team member who gathered information, coordinated with FFID personnel, attended data collection events, and procured data for research purposes.

Research Context

In 2007 and 2008, shortly after being created at Fort Bliss, FFID was responsible for organizing and administering a series of exercises conducted by the 5th Brigade Combat Team, 1st Armored Division, Army Evaluation Task Force (AETF) to integrate, test, and evaluate FCS SO technologies. These SO technologies included (but were not limited to) Urban-Unattended Ground Sensors (U-UGS), Tactical-Unattended Ground Sensors (T-UGS), and the Non-Line of Sight Launch System (NLOS-LS). The 5/1 AD (AETF) was task organized as a Heavy Brigade Combat team. The SO test unit--A Company 2nd Combined Arms Battalion--consisted of one Mechanized Team (2 + 2) with attachments, an engineer platoon, a scout platoon, a COLT team, and a NLOS-LS section.

The FCS integration exercises conducted by 5/1 AD (AETF) followed a rough crawl-walk-run progression from initial desktop computer-based training to leader team constructive simulation training to unit individual and collective live training and, finally, test events. Over the course of the exercises, Soldiers and leader teams were first exposed to the SO technologies, then gained familiarity with the equipment and its usage in operations, then used the equipment to conduct a variety of collective training missions (e.g., attack, defense, and screen). Exercise results led to revisions in the SO tasks themselves as FFID personnel and FCS contractors gained information about equipment usage and trouble spots.

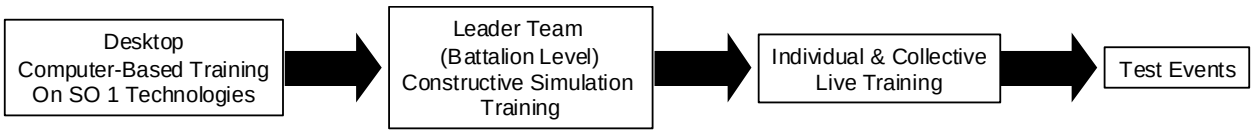


Figure 3. Progression of training, test, and evaluation exercises conducted by the 5/1 AD (AETF).

Observer/Trainers (O/Ts) from FFID informally assessed performance in the initial train-up exercises and conducted more formal assessments during live evaluation exercises and test events. The primary function of O/Ts was to assist in equipment integration and training such that by the end of the last training exercise and before test events all units of the 2nd battalion were working productively with SO technologies.

It is within this context that we designed our research approach and procedure and collected data. The scope of our effort necessitated that we conduct the lion's share of our investigations by integrating our efforts with those of FFID, working with the tasks they were training, sampling from the performance data they collected, and using their observations about how exercise events unfolded. This approach contrasted somewhat with that taken in previous ARI skill retention research in which separate data collection events were held, performance was rated by a group of trained experimenters, and task conditions were relatively controlled (e.g., Rose, et al., 1985; Sanders, 1999). The context in which we operated, necessitated by a wartime environment, was characterized by a wealth of information about SO tasks and their influence on unit performance, but also by competing requirements for performance data. As is documented in this report, much was accomplished with regard to our research goals. Our lessons learned also extend from this context, illuminating the particular challenges of integrated training and

evaluation research programs and the tailored approach necessary for making the most out of available resources.

Skill Retention Research

Background

Previous ARI skill retention research sought to identify factors that influence the rate of individual skill decay (Rose, et al., 1985; Wisher et al., 1991). If tasks could be reliably described according to these factors it would then be possible to examine, using actual performance data, how particular combinations of factors influence performance at some time interval (e.g., 2 months) after initial task proficiency had been obtained. A quantitative relationship between a task's factors and its associated performance data at a particular time interval could serve as a generalizable model of skill decay. Projections well into the future (i.e., beyond the specific time interval studied) could be accomplished by using a theoretical curve to describe the path of skill decay.

Among the challenges faced by the pioneering ARI researchers were identifying the correct factors (i.e., the characteristics of tasks, people, and performance context that actually influence retention), reliably measuring these factors, collecting performance data with the quality and quantity sufficient for modeling, and positing the correct theoretical curve to describe trends in performance over time intervals not studied. Rising to these challenges, the ARI researchers produced the User's Decision Aid (UDA) (Rose, et al., 1985; Rose, et al., 1985a, 1985b). The purpose of the UDA was to provide a means for reliably assessing the characteristics of individual, procedural military tasks. The UDA was a survey that listed 10 task characteristics (e.g., number of performance steps, use of job aids, etc.) which respondents (subject matter experts) used to rate a set of given tasks. For instance, with Macpherson et al. (1989), respondents could use the UDA to rate wheeled vehicle maintenance tasks. The UDA has also been used to rate mobile subscriber equipment operator skills and cannon crewman (field artillery) Skill Level One tasks (Rose, et al., 1985; Sabol et al., 1990) using a variety of methods from survey to on-on-one interviews.

In one series of investigations, researchers examined the correspondence between rated task characteristics to actual performance data at multiple time intervals (two, five, and seven months; Rose, et al., 1985). The results from the final phase of this research indicated that inter-rater reliability using the UDA was generally high and that the UDA predicted retention fairly well for the 22 selected field artillery tasks. Correlations between predicted and actual performance at the 2-month retention interval fell in the .80-.90 range and in the .60-.70 range for 5- and 7-month retention intervals.

The above-described work represents a relatively isolated application of skill theory to account for skill decay when scheduling training. The UDA was validated against actual retention data only a single time, and is not currently in use for scheduling training. A handful of studies of military skill retention have followed this initial research (e.g., Adams, Webb, Angel, & Bryant, 2003; Goodwin, 2006; Sanders, 1999; Stothard & Nicholson, 2001), but a more effective, more widely adopted model than the UDA has not been produced. Technological

advances since the initial research make it trivial develop a sophisticated computerized version of the UDA, simplifying its use and increasing the likelihood that unit trainers would adopt it. However, theoretical advances and changes in the contemporary operating environment make it necessary to revise the UDA itself.

The Present Research

The challenges we faced in our skill retention research mirrored those of the original ARI researchers. Our first job was to re-examine the task characteristics listed in the UDA and revise them to predict retention on FCS-enabled individual tasks. This required the application of theoretical advances in cognitive skill development and retention as well as an analysis of future combat tasks. We also generated a list of characteristics that potentially influence the retention of FCS-enabled collective tasks. We combined these characteristics into a survey analogous to the UDA, which we called the Trainer's Decision Aid (TDA). Separate versions of the TDA survey were created to assess individual and collective task factors.

Our next task was to use the TDA to assess the factors of a set of FCS SO individual and collective tasks, then collect performance data on these tasks. Our goal, analogous to that of the Rose, et al. (1985), was to examine how well the factors we selected predicted skill decay at a particular time interval following the attainment of task proficiency. We would use our data to build and validate a generalized model of FCS skill retention at that time interval and then apply theoretical curves to project retention out into the future.

Revising the User's Decision Aid

Revisions to the UDA were made to address known limitations of the UDA and to extend the UDA in response to recent reviews of skill retention research. Other enhancements to the UDA were made to improve the usability of the survey and to make the survey applicable to both individual and collective FCS tasks. These revisions are documented below. The complete TDA (i.e., revised UDA) for individual tasks is shown in Appendix A. The complete TDA for collective tasks is shown in Appendix B.

Addressing Limitations of the UDA

Predicting performance over time. Although the UDA was demonstrably effective in predicting the pattern of retention across selected field artillery tasks, there was one important way in which its predictions systematically differed from actual performance data: the UDA generally over-predicted skill decay at each retention interval studied (see also Stothard & Nicholson, 2001). Even at a retention interval of two months, the UDA predicted on average a smaller percentage of Soldiers would be "GO" than was borne out by the actual performance data (Rose, et al., 1985). The average difference between predicted and actual performance at two months was relatively small, only five percentage points, but for specific tasks, this difference was as great as 30 percentage points. Such a difference could have substantial practical significance to a commander or unit trainer making judgments about when to schedule refresher exercises.

In addition, the aggregated actual performance data at two, five, and seven months in Rose, et al. (1985) produced a U-shaped retention curve, whereas the UDA predicted a negatively accelerating trend in skill level decreases as time without practice increased (see Figure 4 for schematic depictions of the actual and predicted trends). Large deviations of predicted retention from actual retention at seven months (38 percentage points on average) stemmed from this difference in projected versus actual trends. Based on relatively current reviews of the skill retention literature (Stothard & Nicholson, 2001), however, it was correct to project a nonlinear, exponential rate of skill decay. The unexpected U-shape found in Rose, et al. (1985) may stem from the fact that the retention intervals differed across participants and some re-learning occurred during retention testing.

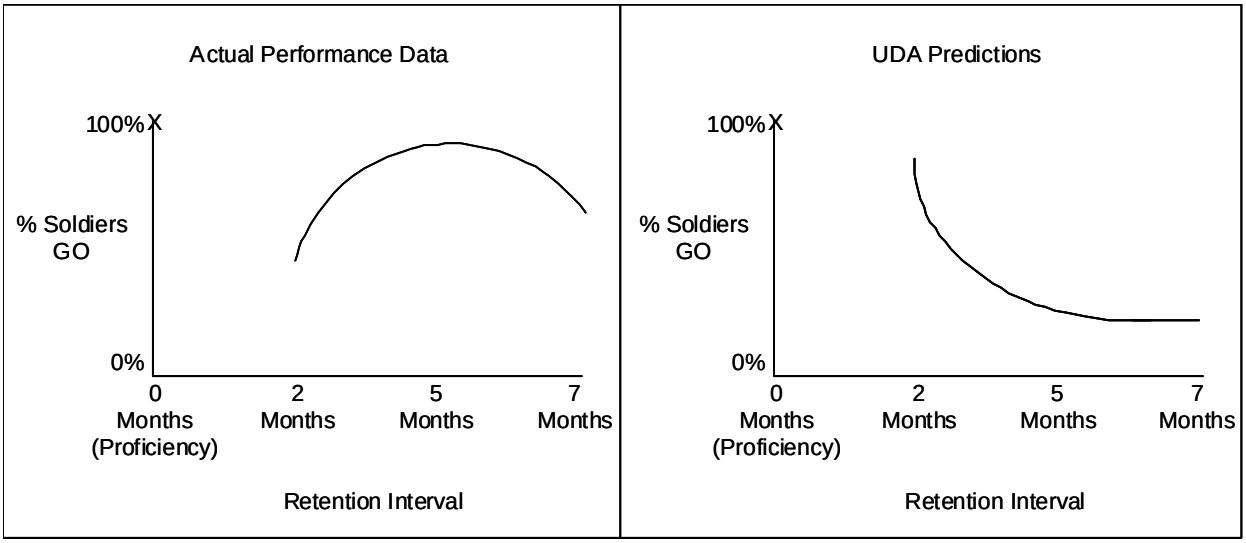


Figure 4. Schematic depiction of actual and predicted performance trends (Rose, Czarnolewski et al., (1985).

For instance, at the 5-month retention interval, just over half of the participants had been tested at two months and half had not. The share of participants who had been tested at two months received an additional opportunity to perform the tasks, thus enabling spaced practice and a reduction in retention interval from five months to three months. Similarly, approximately 75% of the participants tested at seven months had the opportunity to perform the tasks during 5-month testing, reducing their retention interval to two months. Close to 60% of these participants also were tested at the 2-month interval, thus receiving two additional opportunities to practice the tasks. Stothard and Nicholson (2001) have argued that the UDA failed to take into account the effects of practice and re-learning. Their argument is plausible, given the performance data presented in Rose, et al. (1985). It also reflects the reality of the execution and maintenance of skills for active duty service members.

The important implication of this analysis is that a revised UDA should take into account the fact that opportunities to perform tasks on the job are common in the Active Component and that some tasks are practiced more than others. This is true for both individual and collective tasks. Tasks that are performed more than others may show less skill decay at equal retention

intervals than tasks performed less frequently. Such tasks may also produce a shallower downward curve in performance over time. Our revisions of the UDA explicitly ask users to rate the frequency with which tasks are performed on the job. Another important implication of this analysis is that the retention intervals used to validate the revised UDA should be true retention intervals such that practice effects or re-learning do not reduce the interpretability of the data.

Types of tasks evaluated. Rose, et al., (1985b) admonish users of the UDA that certain types of tasks would be difficult to rate. The task types they listed were collective tasks and tasks performed under varying conditions (e.g., night/day). Stothard and Nicholson (2001) further observed that the UDA focused primarily on procedural, as opposed to cognitive, tasks. Many of the mission essential tasks that modern and future Soldiers must carry out have these characteristics; they are cognitively demanding, are usually not procedural in nature, are performed under a wide variety of conditions, and are collective. A revised UDA must be able to account for tasks with these characteristics as well as procedural tasks to provide useful performance predictions for the full spectrum of mission essential tasks. We revised the UDA in this way by adding survey questions that ask users to rate the mental demand of individual tasks (where possible ratings range from simple physical tasks to complex cognitive tasks) and the complexity of mental demand of collective tasks. Also, when asked to rate the frequency with which individual and collective tasks are performed (see previous section), TDA users are prompted to consider the degree to which performance conditions differ. When projecting retention, we also considered the possibility that different types of tasks rated by the TDA may require different shapes of curve to best fit retention data depending, for instance, on whether the skill can be automatized (Stothard & Nicholson, 2001).

Initial skill level. In their review of the UDA, Stothard and Nicholson (2001) noted that the level of initial training must be taken into account to effectively predict retention. In their review of factors influencing skill decay, Arthur, Bennett, Stanush, and McNelly (1998) noted that both the quality and quantity of initial training had a strong influence on mastery, transfer, and retention, particularly for cognitive skills. They further argued that skills that are taught in a manner that is more representative and consistent with the operational environment are more likely to be retained and more likely to transfer to at least similar conditions. This is important for both individual and collective tasks. In both the individual and collective versions of the TDA, there is a question that asks users to rate the degree to which initial training (e.g., schoolhouse education, unit training, etc.) and subsequent practice of the task is reflective of the operational environment. The TDA also asks users to rate the degree of command emphasis on task performance. It was expected that ratings of command emphasis would provide additional information on training quality and likelihood of retention that TDA users would easily relate to (Michalak, 1981).

Extending the UDA

Collective skill retention. The UDA was explicitly designed to predict retention on individual, procedural tasks. Indeed, even now, relatively little is documented about the factors that influence collective skill retention. One exception is an in-depth review report on the topic written by Adams et al., (2003). Like the ARI researchers who produced the UDA, Adams et al.

sought to assist with scheduling military training, but their focus was specifically on collective skills.

Among the challenges they encountered to identifying factors that influence collective skill retention was defining "collective." Put in other words, it is unclear what echelon should be considered the primary unit of analysis. Also, it is unknown how great a proportion of a collective must change before the collective should be considered as having a different identity as before. If the collective is large, turnover of one or two service members may not constitute a shift in the team's identity, particularly if team coordination processes are largely dictated by standard operating procedures (SOP) or if individual roles on the team are not highly interdependent. In contrast, if the collective is small and highly cohesive, the turnover of just one Soldier could significantly change the team's basic structure and function. For both large and small collectives, turnover of key leaders could constitute major "memory loss" for the collective, with complete, or near-complete, re-training necessary upon the assumption of new leadership.

There was also the challenge of defining "collective task" (Adams et al., 2003). As is evident upon review of a training and evaluation outline of a collective task, such tasks comprise a set of sub-tasks and usually sub-sub-tasks. Some of these sub-tasks or sub-sub-tasks may be performed more frequently than the overarching task, making it difficult to determine what level of analysis should be used when predicting retention and scheduling training. Collective tasks differ from individual tasks on at least one important dimension: the critical importance of interpersonal or inter-collective coordination, depending on the size of the collective. Thus, a collective skill may be defined as a skill whose performance requires the contribution of two or more personnel acting interdependently.

Another challenge was defining "retention." Technically, a retention interval is a period of time of no practice on the skill of interest in between performances of the task. In retention studies, the retention interval studied usually is the time between attaining proficiency on the task and its first performance after no practice. For many military collective tasks, however, it is arguable whether proficiency is ever obtained (Adams et al., 2003). This is partly due to the fact that opportunities to train collectively are extremely limited. In addition, the conditions under which collective tasks are performed can vary so widely that it is arguable whether retention is demonstrated in a post-retention-interval performance or transfer of training. Finally, as described above, some components of collective tasks may have been performed during the hypothetical no-practice interval of the overarching task.

A final challenge is determining what curve to use to project collective skill retention into the future. For example, if proficiency on collective tasks is difficult to obtain and the conditions under which the task is performed vary widely, performance may remain consistently at a sub-optimal level. As another example, if the collective is small, highly cohesive, and not subject to frequent turnover, retention may mirror the negatively accelerating exponential curve typical of individual skills.

In light of these challenges, Adams et al., (2003) identified the following categories of factors as potentially influencing collective skill retention:

- Task characteristics (e.g., number of sub-tasks, interdependence of sub-tasks, need for information exchange, etc.).
- Features of the collective (e.g., size, turnover, social environment etc.).
- Characteristics of individuals forming the collective (e.g., aptitude, diversity of experience, etc.).
- Training features (e.g., level of initial training, collective training opportunities, etc.).

We took these categories of factors into consideration when designing the TDA, combining factors where they overlapped and ensuring that factors unique to collective skills (e.g., role interdependence, size of the collective, etc.) were included. Several of these factors also influence the retention of individual tasks (e.g., level of initial training), and have been discussed previously in this report.

Improving UDA Usability and Applicability

Future combat tasks. The UDA survey took into account the important task characteristic of job aids. Rose, et al., (1985) found that job aids could significantly reduce memory demands and enhance retention by externalizing the performance steps in the form of mnemonics, handbooks, SOP, and technology design (i.e., forcing functions that require steps to be taken in a certain order). Technology design is especially critical to the retention of FCS tasks because technology use is their defining characteristic. In his investigation of digital skill retention, Sanders (1999) found that technology usability influenced errors and reduced retention on two critical digital skills of the future force (overlay generation and message sending). As was illustrated by Sanders' study, technology design is critical not only to the performance of individual tasks; the communications necessary for performing collective tasks are digitally mediated, especially at the battalion and higher echelons.

In our revisions of the UDA, both the individual and collective task surveys, we included additional items to account for the impact of technology on performance. We also included an item in the TDA for collective tasks that asks about the use of information management SOP. To the extent that such SOP are developed and followed, memory of what information to share, with whom, and when is externalized and not a cognitive burden to members of the collective. Similarly, standardized methods for handling shared information (e.g., file naming conventions, file organization/location, etc.) could perform the role of a job aid, reducing the cognitive demand of collective tasks.

Reading demand. Although Stothard and Nicholson (2001) praise the usability of the UDA, a review of the survey reveals that it requires a substantial amount of reading on the part of the unit trainer and uses some terminology more familiar to behavioral scientists than military trainers. Understandably, the UDA provides detailed definitions for the various response options associated with each survey question. If read and processed, these definitions could improve the inter-rater reliability of the survey and increase its predictive validity. The downside is that the reading demands of the survey reduce the likelihood that unit trainers, already pressed for time, will use it. Through the collaborative effort of research psychologists and subject matter experts on our team, we designed the TDA questions to involve minimal text. Both the questions and response options were worded using terminology and analogies that military users could be

expected to grasp readily. Where it was possible to use military correlates to behavioral science constructs, these were employed to further simplify the survey.

The Trainer's Decision Aid

Table 1 summarizes the questions in both the UDA and the individual and collective TDA surveys. The complete TDA for individual tasks is shown in Appendix A. The complete TDA for collective tasks is shown in Appendix B. As can be seen in the table, both versions of the TDA borrow liberally from the already effective UDA, but extend the UDA as described above. In contrast to the UDA, both TDA surveys address factors that influence retention above and beyond task characteristics. These additional categories of factors relate to characteristics of the unit and of external influences on performance. The TDA for collective tasks asks the user to relate characteristics unique to collective skill (e.g., role interdependence, size of the collective). It is likely that the number of questions in validated versions of the TDA will be fewer than what is listed below.

Spin Out Tasks Selected

To ensure alignment between the tasks rated using the TDA and the tasks for which performance data would be available, we requested from FFID a list of the SO tasks they intended to evaluate in their 2007-2008 series of evaluation exercises. Our intent was for O/Ts to rate a representative sample of these tasks such that we could ensure the stability and generalizability of our skill retention models. Our modeling effort required that that we select one sample of individual and collective tasks to build the model and a second sample of tasks to validate it.

A complete listing of the SO tasks we selected is shown in Appendix B. As is shown in the appendix, the tasks fell into four categories: Collective – Assault, Individual – UGS, Individual – NLOS, and Collective – NLOS. Collective – Assault tasks were well established, doctrinal collective tasks (e.g., cordon and search, raid) that were modified to include SO technologies in the conditions, standards, and task steps/performance measures. Contrasting with the previous ARI studies of skill retention, the individual and collective skills specific to SO technologies that we examined (UGS, NLOS) were not yet fully established and documented in doctrinal training manuals. We received multiple, differing versions of the NLOS and UGS tasks and the nature of many of these tasks changed, sometimes substantively, as part of the natural course of the technology integration mission carried out by FFID.

Table 1

Comparison of Items on the User's Decision Aid (UDA) and the Trainer's Decision Aid (TDA)

UDA	TDA
Individual Tasks	
1. Are job or memory aids used by the Soldier in performing (and in the performance evaluation of) this task?	1. How much do memory aids reduce the memory demands of this task?
2. How would you rate the quality of the job or memory aid?	2. How many performance steps is the task divided into?
3. Into how many steps has the task been divided?	3. Are the steps in the task required to be performed in a definite sequence?
4. Are the steps in the task required to be performed in a definite sequence?	4. How complex are the mental demands of this task?
5. Does the task provide built-in feedback so that you can tell if you are doing each step correctly?	5. How many facts, terms, names, rules, and/or ideas must a Soldier or leader memorize in order earn a "GO" on this task?
6. Does the task or part of the task have a time limit for its completion?	6. How difficult are the facts, terms, rules and/or ideas that must be remembered?
7. How difficult are the mental processing requirements of this task?	7. How severe is the time pressure under which this task must be performed?
8. How many facts, terms, names, rules, or ideas must a Soldier memorize to do the task?	8. How strongly has the Battalion Commander emphasized the importance of learning this skill?
9. How hard are the facts, terms that must be remembered?	9. How closely did the initial training conditions of this skill match its performance conditions under combat conditions?
10. What are the motor control demands of the task?	10. How often has this skill been used?
	11. On average, how user friendly are the information displays that are <u>most important</u> to performing this task?
	12. How frequently has the technology involved in performing this task changed?
	13. How reliable is the technology involved in performing this task?
	14. What level of information overload do the Soldiers performing this task typically operate under?

(*Table Continues*)

UDA	TDA
	Collective Tasks
N/A – The UDA did not assess the retention factors of collective tasks.	1. What is the size of the collective that this task applies to? 2. How complex are the mental demands of this task? 3. How interdependent are the roles of the people who perform this task? 4. How severe is the time pressure under which this task must be performed? 5. How strongly has the Battalion Commander emphasized the importance of learning this skill? 6. How effective was the training on this collective task? 7. How often has this task (or similar tasks) been performed by the collective? 8. How much have the SOP applicable to this task reduced the difficulty of coordination? 9. On average, how user friendly are the information displays that are <u>most important</u> to performing this task? 10. How frequently has the technology involved in performing this task changed? 11. How reliable is the technology involved in performing this task? 12. What level of information overload does the leadership or overall collective performing this task typically operate under?

Procedure

TDA Administration

Administering the TDA was a two-step process. In the first step, our on-site military subject matter expert provided instruction on how to use the TDA. This instruction was lecture-based with PowerPoint slides, lasted approximately one hour, and was conducted during the O/T Academy held by FFID. The purpose of the O/T Academy was to bring O/Ts up to speed on the performance assessments and instructional role they were to play during the upcoming technology integration exercises. The TDA instruction provided some background on the skill retention research program and involved worked examples of how to rate tasks using the TDA.

In the second step, the TDA was delivered electronically to O/Ts along with other rating materials: an answer sheet and a list of tasks to rate (see Appendix B). The list of tasks included the conditions, standards, and task steps/performance measures for each task in a format resembling a training and evaluation outline. To manage O/T workload and ensure quality ratings, we made three separate task lists, one for UGS, one for NLOS, and one for Collective – Assault, and task lists were assigned only to O/Ts who were designated as experts by their leadership.

We sought multiple raters for each task category such that we could explore the inter-rater reliability of the TDA items (analogous to Rose, et al., 1985) and could form robust ratings of the characteristics of each task. Twenty-nine O/Ts rated the Collective – Assault tasks, 6 O/Ts rated UGS tasks, and 3 O/Ts rated NLOS tasks.

Performance Data Collection

For the purposes of studying skill retention, the final phase of SO technology integration—live exercise—was the optimal time to collect performance data. Given the crawl-walk-run progression of exercises and the ongoing revisions of SO tasks, performance during the final phase of evaluation was the most likely to represent the first opportunity for 5/1 AD (AETF) to conduct tasks after having achieved asymptotic levels of learning. Task proficiency followed by an "empty" retention interval was necessary for our research because we sought to model what happens to skill in the absence of performance opportunities. For this reason, we requested from FFID the "GO/NOGO" and "Trained, Partially Trained, or Untrained (TPU)" ratings assigned to task performance during live exercises.

Results

TDA Rating Data

The TDA rating data for 22 of the 81 selected individual and collective tasks were analyzed. All of these 22 tasks were Collective – Assault tasks. Seven Collective – Assault tasks provided to O/Ts were not rated. The remaining Collective – Assault, UGS, and NLOS tasks not included in the present analyses received ratings that suggested the O/Ts did not fully understand how to use the TDA.

Inter-rater Reliability

On the following task factors (i.e., TDA questions), 27 of 29 raters were in 100% agreement and gave all tasks the same rating:

- 1.1 – Size of Unit (Answer: Platoon).
- 2.4 – SOP (Answer: SOP *significantly reduced* the difficulty of coordinating) – One rater gave one task a different rating.
- 3.1 – User friendliness of displays (Answer: Largely usable).
- 3.2 – How frequently technology changes (Answer: Once a year).
- 3.3 – Technology reliability (Answer: Somewhat reliable).

The two raters not in complete agreement rated some squad-level tasks, plus a subset of the platoon tasks rated by the other raters. For unknown reasons, they assigned ratings differently than the rest of the group, even on platoon tasks. With the exception of two items (1.2 and 1.4, task complexity and role interdependence, respectively), lack of 100% agreement was due to these raters, and their intercorrelations with other raters were generally below .30. Perhaps noteworthy is the fact that these raters were the first two to submit their ratings and generally used more of the scale to make their responses. The remaining data arrived in a single wave several months after the TDA training, suggesting the possibility that ratings were hastily assigned, ironically, with flagging retention of TDA rating skill. Alternatively, the items themselves may have targeted characteristics that truly are uniform across tasks. The fact that "platoon" was the only echelon for which tasks were rated introduces this as a possibility worth noting.

The inter-rater reliabilities (calculated using Cronbach's alpha) of the remaining seven survey questions are shown in the Table 2. Overall, raters appeared to be in very high agreement. High reliability coefficients appeared to stem, however, from nearly uniform ratings across tasks for each item. Again, the exact reason for this uniformity is unknown, but may be due to rater haste, difficulty using the TDA, or the fact that the characteristics rated were in fact largely uniform across tasks. Raters appeared to be very confident in their ratings. On only a handful of occasions across all tasks and TDA items did raters indicate that they were "somewhat confident" instead of "totally confident" in their ratings.

Table 2

Inter-rater Reliabilities on Selected TDA Questions

Question #	Question Description	Reliability
1.2	Complexity of mental demands	.98
1.3	Interdependence of roles	.99
1.4	Severity of time pressure	.97
2.1	Strength of command emphasis	.79
2.2	Effectiveness of training	.99
2.3	How often performed	.99
3.4	Level of information overload	.99

Inter-item Correlations

To conduct inter-item correlations, the modal rating for each task factor (i.e., TDA question) was used. The intercorrelations among five factors (1.1, 2.4, 3.1, 3.2, and 3.3) could not be analyzed because the modal rating for every task was the same, reducing the variance to zero. Correlations among the remaining seven factors ranged from -.34 to .46, indicating that no factor was functionally equivalent to the others. Twelve of 21 correlations were within the bounds of ± .30. Some of the low correlations may in fact be due to the relatively low amount of variance to be accounted for. For instance, for questions 1.4 (time pressure) and 2.1 (command emphasis), the modal factor ratings for all but one or two tasks were the same. On only one TDA question (1.2, mental complexity) did the modal rating across tasks represent the full range of possible values. Some counterintuitive patterns appeared among the correlations, likely due to range restriction issues. For instance, “command emphasis” was uncorrelated with “training effectiveness” and “frequency of performance.”

Alignment between Tasks Rated with TDA and Tasks Performed in Live Exercises

There were overlapping TDA data and live exercise performance data for just twelve tasks, due to a combination of factors. First, as described above, the TDA data were sufficient for only 22 Collective – Assault tasks and no UGS or NLOS tasks. Second, of the Collective – Assault tasks conducted during the live exercise, only 12 produced ratings on more than one platoon (N = 2-6 platoons). Ratings for multiple collectives were necessary to have sufficient variance to build and validate the models. All platoons in 9 of the 12 tasks received the same performance ratings. The small number of tasks for which performance data overlapped with TDA data plus the questionable validity of the TDA data precluded further data analysis.

Lessons Learned

Selecting Tasks

The FFID’s mission is fundamentally one of supporting Army transformation through the preparation of FCS technologies for fielding. For this reason, tasks, both individual and

collective, are subject to change, at times substantively, during the course of the training exercises that lead up to test events. It is therefore important to delay selecting tasks until shortly before TDA task ratings are required. Other factors play a role in when tasks are rated, chiefly O/T availability. However, to the extent that rating can be delayed to ensure that the final version of each task is being used, the research product will be enhanced. In addition, methods for modeling the performance data that will allow for missing tasks (i.e., those tasks that change substantively between being rated and being performed in a live exercise) should be employed where feasible.

Administering the TDA

The questionable quality of the TDA data suggests that more effective training on how to use the TDA is required. This training would involve a better explanation of the skill retention research purpose and procedures using analogies with which the raters are familiar. Observations during the TDA instruction suggested that O/Ts did not fully grasp why they were being asked to rate tasks and how their ratings would be used. More effective training should be followed immediately by conducting task ratings. Experimenters should facilitate the assignment of ratings in one-on-one interview sessions, as was done in Sabol, et al., (1990) or in small group sessions. Administering the TDA in this way would ensure that raters have a thorough understanding of their task and would give experimenters direct visibility on the rating process.

Collecting Performance Data

Collecting sufficient quality performance data for modeling purposes proved to be exceedingly difficult, even with cooperation and facilitation from FFID. First, it is always difficult to conduct research where the unit of analysis is the collective because so much more data are needed, yet data collection is simultaneously more difficult to coordinate. In our research, the only echelons for which more than one collective was assessed during live exercises were platoon and squad.

Another reason that it was difficult to collect usable data was because the purposes of our skill retention research and of FFID's technology integration mission were somewhat at odds with each other. That is, FFID sought to identify the training and technology integration interventions that would optimize performance and reduce variability, whereas we sought to identify the factors that produce decay in performance and increase variability. The O/Ts' use of binary ratings ("GO/NOGO") to characterize unit performance further reduced potential variability. As a result, the data we had to work with showed minimal variability, making it difficult to produce stable models of retention.

Collective Skill Measurement Research

Currently, training performance measures used to assess collective tasks employ either a binary "GO/NOGO" distinction for rating task steps or a summary "TPU" rating for the task as a whole. These broad distinctions may not reflect the actual variance in performance, particularly for complex collective skills. In other words, among collectives rated as "Trained" some

collectives may be “more Trained” than others. Similarly, a collective may be “more GO” on some performance steps than on other performance steps also rated “GO.”

Reducing variance in this way may simplify readiness assessment and reporting, but it simultaneously makes other examinations of readiness more difficult. Chiefly, broad measures may be insensitive to collective performance changes due to skill decay or due to new equipment, updated information displays, personnel changes, advanced training technologies, and other interventions. Gradations in training performance not captured by broad performance measures may also bear an important relation to operational success and could be important targets for training or other interventions. For instance, Evans and Baus (2006) found that a performance aid for conducting troop-leading procedures (TLP) enhanced TLP performance by small-unit leaders, especially during the task step of forming a tentative plan. The performance aid provided unit commanders with a detailed checklist of performance criteria to consider when conducting this and other steps of the TLP. In sum, more refined performance measures could inform decision making about how to schedule and improve training and enhance operational readiness.

The purpose of our collective skill measurement research was to lay down groundwork for producing refined measures of collective performance by units equipped with FCS SO technologies. To achieve our objective, we researched and developed a measure of FCS-enabled troop-leading procedures (TLP) called the TLP Observer Checklist. Our intent was not to create a technologically sophisticated assessment tool but to identify the competencies that should be measured, to specify a feasible observer-based assessment procedure, and to explore the human factors that influence the adoption of refined performance measures. In this section we describe the development and initial administration of the TLP Observer Checklist as well as our lessons learned.

Design of the TLP Observer Checklist

Consistent with measures developed in other ARI-funded projects investigating collective performance assessment (Cianciolo & Sanders, 2006; Leibrecht, Lockaby, Perrault, & Meliza, 2004), the TLP Observer Checklist was designed to go a level deeper than the extant assessment standard. Going a level deeper involved identifying the underlying thought processes and cognitive capabilities that enable the effective achievement of doctrinal performance standards. In this way, scores on refined performance measures could be correlated with performance rated according to doctrinal standards even though the two assessment approaches are designed to capture different constructs. Refined performance measures capture multiple levels of process effectiveness whereas broad measures capture the outcomes of these processes.

In the case of FCS-enabled TLP, the extant assessment standard we used was the doctrinal training and evaluation outline. Specifically, we began with the task steps and performance measures from three such outlines (i.e., TLP as conducted by different types of FCS-equipped platoon: mechanized infantry, armor, and reconnaissance). As shown in the Table 3, the task steps in these outlines were nearly identical, so a single outline was created by aggregating all four. Although maintaining situational understanding was not an explicit task

step for two of the three platoon types, it was difficult to imagine that maintaining situational understanding was not a critical performance requirement for these units.

Table 3

Troop Leading Procedure Task Steps for Different Platoons Equipped with Spin Out Technologies

Task Step	Mechanized Infantry Platoon	Armor Platoon	Reconnaissance Platoon
Maintain Situational Understanding	X		
Issue Warning Order	X	X	X
Mission Analysis	X	X	X
Tentative Plan	X	X	X
Initiate Movement	X	X	X
Conduct Recon	X	X	X
Complete Plan	X	X	X
Issue Orders	X	X	X
Supervise Preparations & Refine Order	X	X	X

All three training and evaluation outlines accounted for the adoption of SO technologies in performance sub-steps and notes. Specifically, a small handful of additional sub-sub-steps in each outline listed the use of SO technologies to enable the performance of the task sub-steps. For instance, for TLP as conducted by reconnaissance platoons, a sub-step of Mission Analysis (i.e., conduct mission, enemy, terrain and weather, troops and support available, time available, and civil considerations analysis) includes the following sub-sub-step: "Determine the intent of employment of T-UGS in regards to higher headquarters' [intelligence, surveillance, and reconnaissance] ISR plan." An example note relating to the use of SO equipment is as follows: "Note. Emplacement and recovery of ISR-UGS nodes and gateways should be rehearsed as a battle drill."

Although the technology differs, from the perspective of assessing larger collective task performance, the demands of TLP as conducted by conventional vs. FCS-equipped units appeared similar but perhaps slightly more complex. It seemed safe to conclude that a performance assessment measure of FCS-enabled TLP should be applicable to TLP as conducted by conventional units. To go a level deeper than doctrinal task steps/performance measures, we extended existing observer-based methods for assessing tactical cognitive expertise (Phillips, Ross, & Shadrick, 2006).

Tactical Thinking Behaviorally Anchored Rating Scales (T-BARS)

The Tactical Thinking Behaviorally Anchored Rating Scales (T-BARS) (Phillips et al., 2006) is an observer-based checklist for assessing the quality of a leader's cognitive processes

underlying decision making during scenario-based training. T-BARS is based on a five-level conceptualization of cognitive expertise: Novice, Advanced Beginner, Competent, Proficient, and Expert. According to this conceptualization, tactical leaders develop cognitively from novice to expert by refining their mental models in the areas of Assets, Mission, Enemy, and Terrain. They move from rigid, rule-based mental models that are insensitive to context toward highly adaptive, rapidly executed, goal-directed mental models.

A key characteristic of the T-BARS is that it provides the observer a set of behavioral targets, or anchors, to look for when assigning ratings of cognitive expertise. The use of behavioral anchors increases the reliability of the measure (Dwyer, Fowlkes, Oser, Salas, & Lane, 1997). In addition, the T-BARS behavioral anchors generalize across scenarios used for training and the mode by which scenarios are presented or carried out (e.g., live training vs. computer simulation vs. tabletop exercise). For instance, in the area of "Enemy," the behavioral anchors for "Novice" include "ignores enemy during mission analysis/planning/execution" and "ignores typical enemy capabilities and assets or states them incorrectly." In contrast, "Expert" behavioral anchors include "articulates how course of action will use terrain, assets, or other resources to deny enemy objective" and "makes a projection about how enemy or populace will react to own actions." At the mid-level of cognitive expertise, "Competent," examples include "generates ideas about what the enemy may be thinking" and "questions how the enemy might respond to own [course of action]."

Extending T-BARS

The T-BARS was designed as a tool for assessing individual decision making. We sought to retain the many state-of-the-art characteristics of the T-BARS (i.e., theoretical basis, behavioral anchors, and generalizability) while simultaneously extending the tool to assess collective functioning demonstrated during planning or collaborative problem solving (see Appendix E, which contains the entire TLP Observer Checklist). Specifically, we modified:

1. *The areas in which level of cognitive development was assessed*. Rather than using the tactical thinking themes of Assets, Mission, Enemy, and Terrain, the areas we selected were the nine task steps listed in our aggregate TLP training and evaluation outline (see Table 3 above). This modification would allow raters to track collective performance in the areas of specific interest to the training organization as well as chronologically throughout the training exercise. Chronological tracking would make the checklist easier to use by eliminating the likelihood that behavioral anchors are spread across checklist pages (see Phillips, et al., 2006). Although the tactical thinking themes certainly would apply to the effective conduct of TLP, our goal was to assess collaborative processes critical to successful *collective* problem solving, including information sharing and shared situation understanding.

2. *The behavioral anchors associated with the five levels of cognitive development*. In the TLP Observer Checklist, the anchors were worded to reflect collaborative activities during planning rather than individual considerations or behaviors involved in tactical decision making. The anchors were based on an analysis of the task sub-steps (or performance measures) listed in the doctrinal training and evaluation outline

and knowledge of common shortfalls in collaborative planning in both military and civilian contexts.

3. *The target application of the checklist*. In contrast to the T-BARS, the TLP Observer Checklist is designed to assess level of cognitive development as reflected in the performance of a specific doctrinal task. The T-BARS may be applied to a large variety of doctrinal tasks, whereas the TLP Observer Checklist is specific to the conduct of TLP.

In sum, the TLP Observer Checklist uses a format similar to that of the T-BARS (see Appendix E). For the nine task steps involved in conducting TLP, behavioral anchors are provided to help observers rate collective performance. The behavioral anchors are generalizable across training scenarios as well as training contexts. Ratings for each behavioral anchor assign one of five levels of expertise. Task step scores are generated by averaging the ratings assigned to the behavioral anchors within the task step. These are then averaged across task steps to produce an overall TLP score. The TLP Observer Checklist differs from the T-BARS in that it is focused on collective as opposed to individual processes and has problem solving as the main task focus as opposed to decision making.

TLP Observer Checklist Development & Administration Process

The format of the TLP Observer Checklist was refined through an iterative development and testing process. First, a rough format of the checklist was applied by an in-house military subject matter expert to observing two computer simulation-based training exercises in which platoons conducted TLP. The initial draft of the TLP checklist was very thorough but had too much information to process and was too cumbersome to use while trying to evaluate the training being conducted. Feedback from the subject matter expert enhanced the design such that it was easier to use (i.e., behavioral anchors were easier to locate in the checklist and observations were easier to mark in the checklist), but also indicated that the behavioral anchors captured the essence of TLP collective performance demands.

A revised version of the checklist was applied by the same subject matter expert to observing three more computer simulation-based training exercises. The revised TLP checklist provided enhanced conditions to evaluate the units while compiling data on a condensed single page format without compromising essential information for accurately and efficiently evaluating ongoing training. It was determined that the revised checklist format was much easier to use, although the data available were insufficient to analyze the psychometric properties of the measure.

TLP Observer Checklist Lessons Learned

Administering the TLP Observer Checklist

Through the iterative development and testing process, it was determined that the TLP Observer Checklist is best administered by someone who has leadership experience in the Army (i.e., and is familiar with TLP) and who is familiar with the scenario to be used for the training

exercise. Military and scenario knowledge is necessary for enhancing the reliability of the checklist by ensuring that behavioral anchors are readily understood and identified within the specific context of the exercise (Dwyer, et al., 1997). Without such knowledge, a rater may misunderstand the behavioral anchors or fail to recognize them as they occur during the training exercise.

Validating the TLP Observer Checklist

To be considered a worthwhile measure of TLP execution, the TLP Observer Checklist must demonstrate the following reliability and validity characteristics (see Cronbach, 1990):

1. That multiple observers with the required expertise (see previous lesson learned) rate different collectives in a common fashion (i.e., inter-rater reliability);
2. That the checklist assesses a representative sample of TLP tasks described in doctrine (i.e., content validity);
3. That the pattern of performance differences among collectives assessed by the checklist is roughly preserved when the same collectives are assessed using independent alternative measures (e.g., Evans & Baus, 2006) of TLP and TLP sub-task execution (i.e., construct validity); and

4. That the pattern of performance differences among collectives assessed by the checklist is roughly preserved when the same collectives are assessed using measures of mission outcomes (i.e., criterion-related validity).

Collecting data that would test these reliability and validity characteristics requires:

- The participation of multiple experts to administer the checklist;
- Access to collectives that vary in terms of their performance on independent alternative measures of TLP and TLP sub-task execution; and
- Access to collectives that vary in terms of their achieved mission outcomes.

We discovered that although it was straightforward to design a TLP Observer Checklist that sampled the range of TLP tasks, collecting the data necessary to validate the measure was much more difficult. First, due to time constraints shouldered by candidate experts, only a single expert (a project experimenter) was available to administer the measure. Second, only the performance of collectives in the simulation phase of training could be assessed by the available expert because planning was distributed and transportation was strictly limited during live training exercises. Third and finally, access to performance outcomes during the simulation phase of training was not possible because O/Ts did not formally assess this phase of training.

Our key lesson learned was that the validation of new performance measures, especially collective performance measures, which involve many more people and much more coordination than individual performance measures, requires extensive socialization of the research objective and procedure to acquire sufficient data. Socialization would achieve the buy-in necessary to enable performance assessment that is off the critical path for the participating training organization and to facilitate coordination, such as transportation across live training events.

Adoption of the TLP Observer Checklist

A significant challenge to socializing collective performance assessment research is the fact that the "GO/NO GO" and "TPU" distinctions are entrenched in Army performance assessment. Unit training status reporting, for example, requires that commanders report the percentage of mission essential tasks that are trained to standard. This percentage is a weighted average of mission essential tasks for which the unit has been rated "T," "P," or "U" (Army Regulation 220-1). Finer distinctions, if made during training assessment, would have to be aggregated into "GO/NOGO" or "TPU" categories to comply with current unit status reporting procedures. Moreover, a binary, pass/fail distinction enhances unit training readiness on paper (i.e., makes it easier to report larger percentages of trained personnel) because a wide range of performance levels could be considered sufficient for effective mission execution in the field. Finally, adopting finer distinctions to characterize "more or less GO" individuals or collectives as official readiness reporting policy requires that data be collected to link varying performance levels in training to varying performance levels in theater. Such data would be extraordinarily difficult to collect due to limited access to theater and the fact that measures of success in a mission context, especially a counterinsurgency mission context, are poorly defined (McGonigle, Casper, Meiman, Cronin, Cronin, & Harris, 2005).

Our key lesson learned with regard to measure adoption is that new measures of collective skill are, for the time being, best used for research purposes. The ability to make fine performance distinctions would inform the enhancement of all aspects of collective performance, including team composition, information technology design, information management procedures, and instructional strategy. Socializing collective performance assessment research therefore involves convincing the leadership of participating training organizations that the research, if not the measures themselves, has value because of its potential to enhance aspects of collective performance of interest to that organization. Ideally, the investigation of new measures of collective performance would be conducted with an eye toward informing the Army's official understanding of readiness and mission success. Demonstrating that fine performance distinctions relate strongly to concrete outcomes of interest to the Army will go a long way toward shaping human resources practices that require reporting such distinctions.

Next Steps

The remainder of our research program will be conducted in concert with FFID's 2009-2010 series of FCS integration exercises. During this time, we will attempt a second data collection to build and validate the TDA and to assess the psychometric properties of the TLP Observer Checklist. Changes to our procedure in this second data collection will be based on the lessons learned in the first year of the research program. These changes are detailed below.

Skill Retention Research

Our next wave of skill retention research will involve a combination of new and previously tested sets of SO tasks. To ensure that the tasks we ask O/Ts to rate are in fact the same tasks that are evaluated during the live FCS integration exercises we will postpone rating using the TDA until the last possible, feasible moment. Given the demands on O/T time due to

the sequencing of exercises, we expect that this moment will occur after the O/Ts have completed their own training but before 5/1 AD (AETF) training on the SO technologies begins in earnest.

To conduct the next set of ratings with the TDA, we will coordinate more closely with FFID to ensure that O/Ts receive quality instruction immediately prior to assigning task ratings. First, better instruction on how to use the TDA will be provided. Enhancements will include an improved explanation for why O/Ts are being asked to provide ratings and clearer instruction on how to use the TDA. Our team's research psychologist will present the training in person and provide answers to questions about the research. Immediately following the training, O/Ts will be asked to provide ratings in small groups, each with a member of our research team present to assist. We will attempt to manage O/T workload by limiting the number of tasks each must rate to 15 or fewer.

Due to the exploratory nature of the FCS integration exercises, it is not expected that the next wave of performance data collection will produce more data with greater variability. We will explore ways to collect additional, more diagnostic performance data through conversations with O/Ts and FFID personnel. If feasible, a protocol will be developed to interview O/Ts immediately after they have observed live exercises to collect additional data on how the units they assessed performed. Based on the task training and evaluation outlines, the protocol will include questions about challenges the unit faced while conducting each performance step. The additional interview data will be used to produce finer gradations in individual and unit performance.

To address the substantial likelihood that tasks will change between the time that they are rated using the TDA and performed during live exercises, we will (1) get specific information on how each task changed to evaluate the severity of the change (i.e., we will produce a method for qualitatively ranking changes and investigating their impact on task identity); (2) explore the use of latent variable modeling techniques so that new tasks may be substituted for severely changed old tasks in our retention models.

The figure below shows a full concept latent variable model of retention that we could use to test the factors that relate to skill decay. The middle and bottom of the figure show a potential data reduction model that could be used to specify the minimum number of factors necessary to account for the variance among the modal O/T ratings. These would be the potential factors influencing skill retention. Note the below model shows only four latent variables, but the number of possible latent variables potentially equals the number of retention factors rated by the TDA (14 for individual tasks, 12 for collective tasks). It is likely, however, that a reduced set of factors will be sufficient to account for the variance in the TDA rating data. New tasks to be substituted for old, severely changed tasks would be rated according to the reduced set of factors and included in the retention analyses.

The middle and top of the figure show a structural equations model linking retention factors to performance during the first and second phase of the Force Development and Test Experiment (FDT&E) (i.e., live evaluation exercises). This model would reveal the relative weights of the retention factors in predicting performance over two retention intervals naturally

occurring during the FFID FCS integration exercises. These weights would then be used to create summary retention scores for each task [consistent with Rose, et al. (1985) each possible rating *within* a retention factor, i.e., each possible response option to a TDA question, would receive an arbitrary weight]. As in Rose, et al. (1985), initial increments in skill decay as a function of increments in summary retention scores would be arbitrarily assigned. Different shapes of retention curve will be fitted to performance data and those with the best fit will be selected to project skill decay beyond the retention intervals studied.

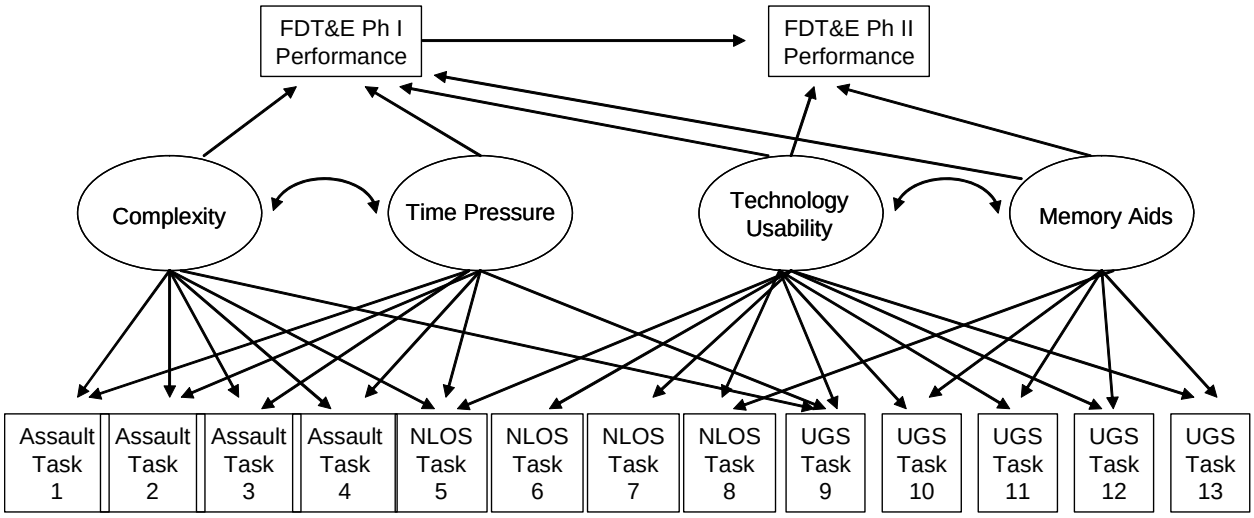


Figure 5. Conceptual latent variable model of skill retention.

It is unlikely that there will be sufficient data to test the full concept model shown above. The data reduction model will have to be tested independently of the structural equations model to have stable parameter estimates.

Collective Skill Measurement Research

Remaining collective skill measurement research will focus on collecting the data necessary to evaluate the psychometric properties of the TLP Observer Checklist. Specifically, with the participation of our on-site subject matter expert, we will identify *a priori* all opportunities to observe TLP in the next set of FCS integration exercises. Given the constraints on observing TLP during live exercises, we will focus our efforts on leader team and small unit training conducted in computer-simulated environments. We will attempt to observe enough collectives so that we can run basic statistical analyses (e.g., *t*-tests and correlations) on the checklist data. Potential barriers to assessing enough collectives will be identified ahead of time, and a plan for working around these barriers will be developed such that the possibility of collecting sufficient data is maximized.

We also will develop a list of variables for which data are required to validate the checklist so that we can assess these variables on our own or request the associated data from FFID. The variables will be constructs with which we expect scores on the checklist to covary, such as deployment experience, unit cohesion, and O/T ratings. We will identify ahead of time which variables FFID collects data on and which variables we will need to collect data on

ourselves. The variables for which we must collect data ourselves will be listed and provided along with the TLP Observer Checklist to our subject matter expert so that we can conduct simultaneous data collection and TLP observations.

It is unlikely that we will have enough raters to conduct inter-rater reliability analyses of the TLP Observer Checklist. With the data available, we will attempt to validate the ability of the measure to distinguish between better and worse performing leader teams as characterized by independent measures of performance (e.g., O/T observations, related mission outcomes, etc.). Given sufficient data we will also investigate various weighting schemes to arrive at an overall TLP score.

Conclusions

Although our efforts to gather and model task performance data were not successful, the present research program represents a step forward in the understanding, assessment, and exploration of collective skill retention. It builds on commonly recognized best practice in skill retention research, begun by ARI in the mid-1980's, and was designed to address specific areas where research and development is especially needed. With refined models of collective skill retention, this research program would be the first of its kind in producing useful quantitative analysis of the factors that influence collective skill retention. In addition, through enhanced performance assessment techniques, this work will provide useful methods for conducting psychological research in the context of ongoing training and evaluation exercises necessary to maintain Army readiness. The next year of this research program promises to be challenging, but the planned research program, based on our lessons learned, should overcome many of the hurdles to building useful individual and collective skill retention models for the Army.

References

Adams, B. D., Webb, R. G., Angel, H. A., & Bryant, D. J. (2003). *Development of theories of collective and cognitive skill retention* (CR-2003-078). Toronto, Ontario: Defence Research and Development Canada.

Cianciolo, A. T., & Sanders, W. R. (2006). *Wargaming effectiveness: Its conceptualization and assessment* (Technical Report 1178). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Cronbach, L. J. (1990). *Essentials of psychological testing (Fifth edition)*. New York: Harper Collins.

Dwyer, D. J., Fowlkes, J. E., Oser, R. L., Salas, E., & Lane, N. E. (1997). Team performance measurement in distributed environments: The TARGETs methodology. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory, research, and applications* (pp. 137-153). Mahwah, NJ: Lawrence Erlbaum Associates.

Evans, K. L., & Baus, E. A. (2006). *Improving troop leading procedures at the Joint Readiness Training Center* (Research Report 1852). Arlington, VA. U.S. Army Research Institute for the Behavioral and Social Sciences.

Goodwin, G. A. (2006). *The training, retention, and assessment of digital skills: A review and integration of the literature* (Research Report 1864). Arlington, VA. U.S. Army Research Institute for the Behavioral and Social Sciences.

Leibrecht, B. C., Lockaby, K. J., Perrault, A. M., & Meliza, L. L. (2004). *Digital proficiency levels for the brigade and battalion battle staff* (Research Report 1826). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Macpherson, D., Patterson, C., & Mirabella, A. (1989). *Application of ARI skill retention model to wheel vehicle maintenance tasks* (Research Report 1538). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

McGonigle, T. P., Casper, W. J., Meiman, E. P., Cronin, C. B., Cronin, B. E., & Harris, R. R. (2005). The relationship between personnel support programs and readiness: A model to guide future research. *Military Psychology, 17*(1), 25-39.

Michalak, D. F. (1981). The neglected half of training. *Training and Development Journal, 35* (5), 22-28.

Phillips, J. K., Ross, K. G., & Shadrick, S. B. (2006). *User's guide for tactical thinking behaviorally anchored rating scales* (Research Product 2006-05). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Rose, A. M., Czarnolewski, M., Gragg, F., Austin, S., Ford, P., Doyle, (1985). *Acquisition and retention of soldiering skills* (Research Report 671). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Rose, A. M., Radtke, P. H., Shettel, H. H., & Hagman, J. D. (1985a). *User's manual for predicting military task retention* (Research Product 85-13). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Rose, A. M., Radtke, P. H., Shettel, H. H., & Hagman, J. D. (1985b). *User's manual for predicting military task retention, Volume 2: Revision* (Research Product 85-26). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Sabol, M, Chapell, L., & Meiers, C. (1990). *Predicted decay of mobile subscriber equipment operator skills* (Research Product 90-11). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Sanders, W. (1999). *Digital procedural skills retention for selected M1A2 tank Inter-Vehicular Information System (IVIS) tasks* (Research Report 1773). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Stothard, C., & Nicholson, R. (2001). *Skill Acquisition and Retention in Training: DSTO Support to the Army Ammunition Study* (Report DSTO-CR-0218). Edinburgh, South Australia: Defence Science & Technology Organisation: Electronics and Surveillance Research Laboratory.

Wisher, R., Sabol, M., Sukenik, H., & Kern, R. (1991). *Individual Ready Reserve (IRR) call-up: Skill decay* (Research Report 1595). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Wong, L. (2002). *Stifling innovation: Developing tomorrow's leaders today*. Carlisle Barracks, PA: Strategic Studies Institute, U.S. Army War College.

Appendix A

UDA Example Item

Question 6. Does the task or part of the task have a time limit for its completion?

Answer Choice	Scale Value
• There is no time limit	40
• There is a time limit, but it is fairly easy to meet under test conditions	35
• There is a time limit and it is difficult to meet under test conditions	0

Definitions

The first choice mean s that no time limit has been established for the task or any part of the task, so that a "GO" may be achieved even through one Solider may take much longer to do the task than another Soldier. This choice is also appropriate when a time limit is so liberal that no one ever fails to meet it.

The second choice above applies to those tasks, such as "Assemble the M60 Machinegun," that have a time limit that some Soldiers may find difficult to meet. In this case, the task summary has set a time limit that "pressures" the average Soldier at bit, but only a few would get a "NO GO" because of it.

The third choice is for tasks that have a time limit that is difficult to meet. Safety and combat-related tasks, such as "Sight a Target through the Gunner's Telescope" within 10 seconds would fall into this category. Soldiers being tested on this kind of task often get a "NO GO" on the basis of time alone.

Appendix B

TDA Survey – Individual Tasks

I. INTRODUCTION

The current OPTEMPO and the complexity of the operational environment reduce the time unit trainers have to train an increasing number of tasks. Among the aids that would help unit trainers meet this job demand is assistance in prioritizing training requirements. Effective prioritization enables unit trainers to achieve optimal training impact given the limited time available.

An important factor to consider when prioritizing training requirements is the level of retention that particular skills maintain over various periods of time. A skill that decays more quickly or more thoroughly will require more frequent retraining, particularly if there are limited opportunities to practice that skill on the job.

The Trainer's Decision Aid (TDA) was developed by the Army Research Institute to help unit trainers prioritize training requirements by providing predictions of skill decay over time. The TDA is based on decades of skill retention research.

What is the Trainer's Decision Aid (TDA)?

Basically, the TDA is a survey. It consists of a short set of questions that ask the unit trainer to make judgments about the aspects of an individual task that might affect retention. These aspects include:

1. The characteristics of the task itself (e.g., How many steps does it have?);
2. The way the unit trains the task (e.g., What is the commander's emphasis on the performance of the task?); and
3. External influences on task performance (e.g., How often does the equipment change?)

The survey is divided into three sections—Task Characteristics, Unit Characteristics, and External Factors—to help unit trainers focus on what task aspect is being judged. When a task evaluation is completed, the TDA reports the predicted level of skill retention at several time intervals, ranging from 1 to 20 weeks.[1]

[1] Skill retention past 20 weeks is not predicted by the TDA because skill retention research indicates that skill decay generally levels out within approximately 2-3 months.

What Is A Skill-Retention Prediction?

The TDA makes predictions about the likelihood that skills will be retained after initial training. Based on the judgments a unit trainer makes of a particular individual task, the TDA will report the expected percentage of individuals able to earn a "GO" at increasing intervals of time. An example of how this looks is shown in the table below:

Task: Prepare a T-UGS System for Operation

Weeks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
% GO	98	90	83	76	70	64	59	55	53	52	49	49	49	49	48	48	47	47	46	46

It is important to remember that the purpose of the TDA is to *not* to make command decisions about when or what to train. Instead, retention predictions *inform* decision making by indicating the level of skill decay at certain points in time after initial learning. The commander must decide what percentage of individuals at "GO" represents acceptable readiness.

Who Should Use the TDA?

Subject matter expertise is necessary to ensure that task judgments (and corresponding retention predictions) are as accurate as possible. Task judges should not only be familiar with the task to be evaluated, but also should be very familiar with how it is typically (1) trained; (2) performed during operations; and (3) prioritized as a training and evaluation requirement.

Task judges therefore should be experienced evaluators of the task being rated. For large units, in which subject matter expertise is distributed across warfighting functions, it may be necessary for the S3 or other unit trainer to recruit the participation of additional experts in the task evaluation process. Recruiting the help of others to ensure accuracy is strongly encouraged because accuracy is essential for making effective skill retention predictions.

How Should the Task-Judgment Process be Carried Out?

The task-judgment process has three components:

1. The TDA survey (an electronic word file).
2. The TDA answer sheet (an electronic excel file).
3. A list of individual task descriptions (an electronic word file).

To make the judgment process go as smoothly as possible, the TDA survey should be printed out. The task descriptions and the TDA answer sheet should be opened up on the computer, with the task descriptions full-size in the background and a reduced-size answer sheet in the foreground. Reducing the answer sheet vertically by half allows the user to see the description of the task being judged while at the same time entering judgment data in the answer sheet. The printed TDA makes a handy reference without requiring an additional window to be open on the computer screen.

One task should be evaluated at a time. Although the TDA survey is brief, users should set aside time to make considered judgments about each task of interest, particularly when multiple tasks must be judged. In the case of multiple tasks, judges should set aside blocks of time to focus on making task judgments and allow for short breaks during the judgment process. Doing this should reduce fatigue and enhance concentration on each task judgment.

When task ratings are complete, the electronic answer sheet should be emailed to the point of contact for this research project, SGM (Ret) Jackson. To keep the answer sheets of multiple judges separate, it is helpful to put a last name or initials in the answer sheet filename before sending it to Mr. Jackson.

SECTION I – Task Characteristics

The seven questions in this section ask about the characteristics of the task (e.g., how difficult the task is). In general, these questions ask: "***On average (i.e., across units)***, *what are the implications of this task's characteristics for performance?*" Therefore, task characteristics are to be judged largely independently of knowledge of the individuals to be trained.

Section I, Question 1a. Quality of Memory Aids

How much do memory aids reduce the memory demands of this task?

1. Aspects of the equipment/technology used for this task actually *increase* memory demands.
2. Existing memory aids for this task don't get used/There are no memory aids.
3. Memory aids *somewhat reduce* memory demands on *critical* task components.
4. Memory aids *significantly reduce* memory demands on *critical* task components.
5. Memory aids *totally* eliminate memory demands.

Definitions. *Memory aids* are designed to guide on-the-job performance and to minimize recall. Memory aids may take many forms, such as:

- Acronyms (e.g., SALUTE, OAKOC).
- Technical manuals or pamphlets (e.g., Soldier/leader handbooks, smart cards).
- Labels or instructions printed on or attached to equipment or containers.
- Checklists, flowcharts, worksheets, decision tables, and system-fault tables.
- Written instructions (e.g., on reports or forms).
- Help menus, intuitive interface designs.
- Built-in feedback (i.e., equipment or digital interface doesn't allow steps to be performed out of sequence).
- Standard operating procedures.

Note memory aids must be *used* to reduce memory demands. The typical usage of memory aids should be taken into consideration when answering this question.

Critical task components are those task components that carry the most weight in GO/NO GO ratings. In contrast, superficial task components are included in task descriptions but typically are not observed closely or rated during performance evaluation. For example, the acronym OAKOC may help a leader remember the planning aspects of terrain, but not how to leverage terrain for tactical advantage. The weight of task components in GO/NO GO decisions may vary, depending on the performance rater or task conditions. The characteristics of the most likely performance rater and task conditions should be taken into account when considering which task components are critical.

Intuitive interface designs are easy to use and place minimal requirements on users to train on or remember what interface features do. They also do not require users to override their natural inclinations to use interface features correctly. Unintuitive, hard to use interfaces can actually make a task more difficult to perform without error. Note some users may have more technical savvy than others. When considering the intuitiveness of interface designs, the typical user should be taken into account.

Section I, Question 1b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section I, Question 2a. Number of Steps

How many performance measures is the task divided into?

1. More than 10 steps.
2. 6 to 10 steps.
3. 2 to 5 steps.
4. 1 step.

Definitions. Some task performance measures have task sub-steps listed under them. These should be counted. Do not count performance measures relating to whether the Soldier performed the task steps in sequence or within a certain time period. These measures are viewed only as *scoring criteria* and not as task steps.

If the task summary does not provide sufficient information, or if you feel that a task has not been accurately divided into performance steps, the following guidance may be helpful:

- A *step* is a separate physical or mental activity within a task and has a well-defined, observable beginning and end. A step must be performed to complete a task correctly.

Section I, Question 2b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section I, Question 3a. Sequence Requirements

Are the steps in the task required to be performed in a definite sequence?

1. N/A – Task has only 1 step.
2. Some are and some are not.
3. All are.
4. None are.

Definitions. If a task or parts of it are supposed to be performed in sequence, there must be a statement to that effect in the task summary (e.g., "Do, in order, all steps to clear the object from the casualty's throat."). In the absence of any statement about sequence, assume that sequence for that task is not required, even though there may be a natural or preferred order for doing the steps.

Section I, Question 3b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section I, Question 4a. Mental Demands

How complex are the mental demands of this task?

1. Very complex mental demands.
2. Complex mental demands.
3. Simple mental demands.
4. Almost no mental demands.

Definitions. When making a judgment about complexity, consider the following definitions of mental demand:

1. A task makes *very complex mental demands* if it requires rapid analysis and decisions based on detailed, technical, incomplete information (e.g., planning an attack, troubleshooting complex equipment) and using the input from numerous, diverse resources.
2. A task makes *complex mental demands* if it requires the Soldier to make a choice or decision based on subtle but discrete clues (e.g., setting priorities for fixed targets, identifying different types of aircraft or vehicles) and the cues come from multiple sources.
3. A task makes *simple mental demands* if it involves making gross comparisons (e.g., estimating relative size, weight, or distance; performing simple computations) using relatively few perceptual or data inputs to the comparison process
4. A task makes *almost no mental demands* if it is essentially physical (vice mental), highly repetitive, and/or involves only one or two different direct perceptual inputs (e.g., visual scanning).

Note this question usually cannot be answered entirely on the basis of the task summary, but the correct choice must often be deduced from a careful reading of the summary and first-hand knowledge of the task itself. The complexity of a task should be assessed independently of the skill level of the individual (e.g., regardless of the math student, calculus is more complex than algebra).

Section I, Question 4b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section I, Question 5a. Number of Facts

How many facts, terms, names, rules, and/or ideas must a Soldier or leader memorize in order earn a "GO" on this task?

1. Very many (more than 8).
2. Some (4-8).
3. A few (1-3).
4. None (or memory aids provide all necessary information).

Definitions. This question addresses the number of isolated pieces of information a Soldier must remember to do the task, not the difficulty of remembering them. Examples of the type of information that may have to be remembered include the following:

- Operational terms.
- Battlefield calculus formulas.
- Codes or call numbers.
- Technical names, specifications, or tolerances.
- Doctrinal principles or rules of thumb.

Section I, Question 5b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section I, Question 6a. Difficulty of Facts

How difficult are the facts, terms, rules and/or ideas that must be remembered?

1. Extremely difficult.
2. Somewhat hard.
3. Not at all hard.
4. Not applicable (none to remember or memory aids provide all of the needed information).

Definitions. Facts and terms that have a close connection to the task itself are more likely to be remembered. For example, the terms *firing pin* and *whip antenna* have a logical relationship to their function and are easy to recall. Specific, detailed, or technical information that is unrelated to the task is more difficult to recall. Call signs and radio frequencies are examples of difficult-to-remember information because they are purposely assigned at random but must be used with precision. The average difficulty of facts should be considered, rather than the difficulty depending on the quality of particular individuals.

Section I, Question 6b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section I, Question 7a. Time Limits

How severe is the time pressure under which this task must be performed?

1. Very severe.
2. Somewhat severe.
3. Not at all severe.

Definitions. When making a judgment about time pressure, consider the following definitions of severe:

- Time pressure is *very severe* if there is simply not enough time to perform the task effectively or completely, no matter how skilled the individual.
- Time pressure is *somewhat severe* if there is limited amount of time to perform the task, such that only skilled individuals can complete the task effectively.
- Time pressure is *not at all severe* if there is no time pressure – all individuals have sufficient time to complete the task effectively, regardless of skill level.

Section I, Question 7b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

SECTION II – Unit Characteristics

The three questions in this section ask about the characteristics of the unit as they relate to emphasizing, training, and performing the task of interest (e.g., the quality of initial training of the task). In general, these questions ask: "***In your unit****, how are training and operations conducted for the task in question?*" Therefore, unit characteristics are to be judged using knowledge of the particular unit to be trained and the task of interest.

Section II, Question 1a. Commander's Emphasis

How strongly has the battalion commander emphasized the importance of learning this skill?

1. This skill is not a command priority.
2. This skill receives less than average emphasis.
3. This skill receives average emphasis.
4. This skill receives greater than average emphasis.
5. This skill is a top command priority.

Definitions. *Commander's emphasis* is the level of priority placed on performing a particular task well. Skills that are a top command priority are trained and evaluated most often and most rigorously. Performing these skills well is actively supported and rewarded. In contrast, skills that are not a command priority may be trained, but they are not evaluated (or not evaluated as rigorously) and performing these skills well is not rewarded or emphasized.

Note in some rating conditions it will be unknown what emphasis a particular commander has placed or will place on a particular skill. In these cases, the rating associated with the expected or typical command emphasis should be assigned.

Section II, Question 1b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section II, Question 2a. Initial Training and Performance Conditions

How closely did the initial training conditions of this skill match its performance conditions under combat conditions?

1. The conditions of initial training in no way resembled combat conditions.
2. The conditions of initial training bore some resemblance to combat conditions.
3. Initial training was conducted under roughly similar conditions as combat conditions.
4. Initial training was conducted under the same conditions as under combat conditions.

Definitions. The *match* between initial training conditions and performance conditions corresponds to the "train as you fight" concept. Fully matched training and performance conditions such as might be found in live exercises at combat training centers involve, for example, the same:

- Equipment.
- Stress level.
- A thinking, realistic, unscripted enemy and or non-combatant population.
- Time constraints.
- Range in performance contexts (e.g., day and/or night, supported missions, communications availability, etc.).
- Consequences of errors.

Initial training conducted under *roughly similar* conditions may not involve the same stress level or the full range in performance contexts, but does feature such instruction as hands-on training with actual equipment and scenario-based practical exercise. Complete operational tasks are practiced.

Initial training that bears *some resemblance* to the operational performance conditions does not involve the use of actual equipment or the practice of complete operational tasks. It generally involves some combination of simulation training, demonstrations, and lecture/conference.

Section II, Question 2b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section II, Question 3a. Frequency of Use

How often has this skill been used?

1. Weekly.
2. Monthly.
3. Quarterly.
4. Other (i.e., less frequently than quarterly).
5. Daily.

Note a skill may be used off-duty as well as on-duty if it is performed as part of a self-study requirement.

Section II, Question 3b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

SECTION III – External Factors

The four questions in this section ask about factors outside of the unit's sphere of influence as they relate to conditions and technologies required to perform the task of interest (e.g., the usability of technology). In general, these questions ask: "***In your unit**, how do characteristics of the external environment influence the performance of the task in question?*" Therefore, external factors are to be judged using knowledge that relates to the particular unit to be trained and the task of interest.

Section III, Question 1a. Information Display User Friendliness

On average, how user friendly are the information displays that are most important to performing this task?

1. Difficult to use information displays actually make this task harder.
2. Somewhat usable.
3. Largely usable.
4. Completely user friendly/Not applicable (digital displays are not used at this echelon).

Definitions. *User friendly information displays* are easy to use, not overly complex, and place minimal requirements on users to train on, locate, or remember what interface features do and how they work together. They also do not require users to override their natural inclinations to use interface features correctly. Hard to use interfaces can actually make a task more difficult to perform without error. When judging information displays, consider the overall impact of user friendliness, specifically for those most important to executing the task.

Note that procedures may be developed and used as workarounds to for hard-to-use information displays. These should *not* be considered when answering this question.

Section III, Question 1b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section III, Question 2a. Available Technology

How frequently has the technology involved in performing this task changed?

1. Several times a year.
2. Once a year.
3. Every couple of years.
4. Never/Not applicable (No technology is used to perform this task).

Definitions. When answering this question it is important to consider how often the technology *in the unit* changes. For example, if the technology changes several times a year, but unit receives new technology once a year, then "once a year" should be selected.

Section III, Question 2b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section III, Question 3a. Technology Reliability

How reliable is the technology involved in performing this task?

1. Totally unreliable.
2. Somewhat unreliable.
3. Somewhat reliable.
4. Totally reliable/ Not applicable.

Definitions. The reliability of technology is dependent on a number of factors, including:

- The degree to which the technology performs to government operational readiness specifications.
- The ruggedness of the technology in the conditions typical of field use.
- The degree to which the technology has been tested to ensure its full functionality with other software/equipment and for use in all operations.
- The availability of auxiliary or supporting equipment necessary to make the technology operate effectively.
- The degree to which the technology is practically useful to Soldiers.

Section III, Question 3b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section III, Question 4a. Information Overload

What level of information overload do the Soldiers performing this task typically operate under?

1. High levels of information overload.
2. Moderate levels of information overload.
3. Low levels of information overload.
4. No information overload.

Definitions. *Information overload* pertains to the sheer amount of information pushed to the Soldier at the time of task performance. Information overload reduces the ability to learn, perform, and retain most tasks, and may come from numerous sources, including:

- Multiple command and control information displays that must be used to perform missions.
- Frequent changes to technology and/or tactics.
- Highly complex missions with high-stakes outcomes.

Section III, Question 4b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Appendix C

TDA Survey – Collective Tasks

I. INTRODUCTION

The current OPTEMPO and the complexity of the operational environment reduce the time unit trainers have to train an increasing number of tasks. Among the aids that would help unit trainers meet this job demand is assistance in prioritizing training requirements. Effective prioritization enables unit trainers to achieve optimal training impact given the limited time available.

An important factor to consider when prioritizing training requirements is the level of retention that particular skills maintain over various periods of time. A skill that decays more quickly or more thoroughly will require more frequent retraining, particularly if there are limited opportunities to practice that skill on the job.

The Trainer's Decision Aid (TDA) was produced by the Army Research Institute to help unit trainers prioritize training requirements by providing predictions of skill decay over time. The TDA is based on decades of skill retention research and has been validated using actual skill retention data.

What is the Trainer's Decision Aid (TDA)?

Basically, the TDA is a survey. It consists of a short set of questions that ask the unit trainer to make judgments about the aspects of a collective task that might affect retention. These aspects include:

1. The characteristics of the task itself (e.g., How many steps does it have?);
2. The way the unit trains the task (e.g., What is the commander's emphasis on the performance of the task?); and
3. External influences on task performance (e.g., How often does the equipment change?)

The survey is divided into three sections—Task Characteristics, Unit Characteristics, and External Factors—to help unit trainers focus on what task aspect is being judged. When a task evaluation is completed, the TDA reports the predicted level of skill retention at several time intervals, ranging from 1 to 20 weeks.[2]

[2] Skill retention past 20 weeks is not predicted by the TDA because skill retention research indicates that skill decay generally levels out within approximately 2-3 months.

What Is A Skill-Retention Prediction?

The TDA makes predictions about the likelihood that skills will be retained after initial training. Based on the judgments a unit trainer makes of a particular collective task, the TDA will report the expected likelihood a collective will earn a "GO" rating at increasing intervals of time. An example of how this looks is shown in the table below:

Task: Establish & Maintain NLOS-LS PLT Control Cell

Weeks	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Likelihood GO	98	90	83	76	70	64	59	55	53	52	49	49	49	49	48	48	47	47	46	46

It is important to remember that the purpose of the TDA is to *not* to make command decisions about when or what to train. Instead, retention predictions *inform* decision making by indicating the level of skill decay at certain points in time after initial learning. The commander must decide what likelihood of a GO on a particular collective task represents acceptable readiness.

Who Should Use the TDA?

Subject matter expertise is necessary to ensure that task judgments (and corresponding retention predictions) are as accurate as possible. Task judges should not only be familiar with the task to be evaluated, but also should be very familiar with how it is typically (1) trained; (2) performed during operations; and (3) prioritized as a training and evaluation requirement.

Task judges therefore should be experienced evaluators of the task being rated. For large units, in which subject matter expertise is distributed across warfighting functions, it may be necessary for the S3 or other unit trainer to recruit the participation of additional experts in the task evaluation process. Recruiting the help of others to ensure accuracy is strongly encouraged because accuracy is essential for making effective skill retention predictions.

How Should the Task-Judgment Process be Carried Out?

The task-judgment process has three components:

1. The TDA survey (an electronic word file).
2. The TDA answer sheet (an electronic excel file).
3. A list of individual task descriptions (an electronic word file).

To make the judgment process go as smoothly as possible, the TDA survey should be printed out. The task descriptions and the TDA answer sheet should be opened up on the computer, with the task descriptions full-size in the background and a reduced-size answer sheet in the foreground. Reducing the answer sheet vertically by half allows the user to see the description of the task being judged while at the same time entering judgment data in the answer sheet. The printed TDA makes a handy reference without requiring an additional window to be open on the computer screen.

One task should be evaluated at a time. Although the TDA survey is brief, users should set aside time to make considered judgments about each task of interest, particularly when multiple tasks must be judged. In the case of multiple tasks, judges should set aside blocks of time to focus on making task judgments and allow for short breaks during the judgment process. Taking breaks should reduce fatigue and enhance concentration on each task judgment.

When task ratings are complete, the electronic answer sheet should be emailed to the point of contact for this study, SGM (Ret). To keep the answer sheets of multiple judges separate, it is helpful to put a last name or initials in the answer sheet filename before sending it to Mr. Jackson.

SECTION I – Task Characteristics

The four questions in this section ask about the characteristics of the task (e.g., the complexity of the mental demands it makes). In general, these questions ask: "***On average (i.e., across units)****, what are the implications of this task's characteristics for performance?*" Therefore, task characteristics are to be judged largely independently of knowledge of the particular unit to be trained.

Section I, Question 1a. Size of the Collective

What is the size of the collective that this task applies to?

1. Company/Troop/Battery.
2. Platoon.
3. Section.
4. Squad.
5. Crew.

Definitions. The size of the collective affects the complexity of the task being performed. Tasks performed by larger collectives must be performed under more diverse conditions using a greater number of "moving parts." Collective tasks performed by larger collectives also are more likely to be mediated by digital information displays and there are more sources of information beyond the direct control of the collective.

Note that the same task may apply to multiple collectives (e.g., platoon and squad). If that is the case when answering this question, select the particular collective for which you are trying to schedule training. The remainder of your present judgments for this task should apply to that collective.

Section I, Question 1b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section I, Question 2a. Mental Demands

How complex are the mental demands of this task?

1. Very complex mental demands.
2. Complex mental demands.
3. Simple mental demands.

Definitions. When making a judgment about complexity, consider the following definitions of mental demand:

1. A collective task makes *very complex mental demands* if it;
   - Is typically performed under novel (i.e., untrained in) conditions.
   - Requires rapid decision making that involves the analysis of information coming from multiple diverse sources.
   - Involves unit structures that deviate significantly from doctrinal norms (e.g., integration with host nation security forces, attachment of a dissimilar unit) such that new knowledge is required to work effectively together; and
   - Depends on the successful performance of numerous combined arms supporting tasks.
2. A collective task makes *complex mental demands* if it;
   - Is typically performed under different conditions (e.g., night vs. day).
   - Requires decision making that involves information coming from multiple similar sources.
   - Involves unit structures that deviate slightly from doctrinal norms (e.g., one less platoon; the attachment of a like unit) such that novel modes of coordination are required; and
   - Depends on the successful performance of several combined arms supporting tasks.
3. A collective task makes *simple mental demands* if it;
   - Is routine.
   - Is largely physical (vice mental).
   - Is typically performed under the same or similar conditions.
   - Depends on the successful performance of relatively few supporting tasks.

Note that complexity applies to the mental demands placed on the ***leadership or collective as a whole*** that conducts this task. The complexity of a collective task should be assessed independently of the quality of the unit (e.g., regardless of a group's organization skills, coordinating a multi-day conference is more complex than arranging a dinner party).

Section I, Question 2b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section I, Question 3a. Role Interdependence

How interdependent are the roles of the people who perform this task?

1. Totally interdependent.
2. Largely interdependent.
3. Somewhat interdependent.
4. Not at all interdependent.

Definitions. *Interdependent* roles are non-overlapping (i.e., no two people perform the same activities and the results of one person's work are required for another person to carry out their function). Totally interdependent roles are more commonly found in complex, combined arms collective tasks.

Note that some tasks may have interdependent roles, but each person in the collective can perform the roles of all of the other people, if called upon. In such a case, the roles should not be considered interdependent.

Section I, Question 3b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section I, Question 4a. Time Limits

How severe is the time pressure under which this task must be performed?

1. Very severe.
2. Somewhat severe.
3. Not at all severe.

Definitions. When making a judgment about time pressure, consider the following definitions of severe:

- Time pressure is *very severe* if there is simply not enough time to perform the task effectively or completely, no matter how skilled the collective.
- Time pressure is *somewhat severe* if there is limited amount of time to perform the task, such that only skilled collectives can complete the task effectively.
- Time pressure is *not at all severe* if there is no time pressure – all collectives have sufficient time to complete the task effectively, regardless of skill level.

Section I, Question 4b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

SECTION II – Unit Characteristics

The four questions in this section ask about the characteristics of the unit as they relate to emphasizing, training, and performing the task of interest (e.g., the use of SOP to facilitate information management and coordination). In general, these questions ask: "***In your unit****, how are training and operations conducted for the task in question?*" Therefore, unit characteristics are to be judged using knowledge of the particular unit to be trained and the task of interest.

Section II, Question 1a. Commander's Emphasis

How strongly has the battalion commander emphasized the importance of learning this skill?

1. This skill is not a command priority.
2. This skill receives less than average emphasis.
3. This skill receives average emphasis.
4. This skill receives greater than average emphasis.
5. This skill is a top command priority.

Definitions. *Commander's emphasis* is the level of priority placed on performing a particular task well. Skills that are a top command priority are trained and evaluated most often and most rigorously. Performing these skills well is actively supported and rewarded. In contrast, skills that are not a command priority may be trained, but they are not evaluated (or not evaluated as rigorously) and performing these skills well is not rewarded or emphasized.

Note in some rating conditions it will be unknown what emphasis a particular commander has placed or will place on a particular skill. In these cases, the rating associated with the expected or typical command emphasis should be assigned.

Section II, Question 1b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section II, Question 2a. Training Quality

How effective was the training on this collective task?

1. Not at all effective.
2. Somewhat effective.
3. Largely effective.
4. Extremely effective.

Definitions. *Effective* collective training has the following features:

- All required personnel participated in the training;
- Supporting collective tasks were sufficiently trained beforehand;
- Training used and refined unit SOP;
- Training emphasized collective processes (e.g., information management, shared situation awareness);
- Collective performance was evaluated and feedback provided;
- Training conditions roughly matched (in type and variety) the operational performance conditions.
- Proficiency was achieved by the end of training.
- Necessary resources (equipment, vehicles, opposing forces, etc.) were available.
- Equipment used was the same as what will be used in the field.

Note it is possible that the quality of collective training will be unknown or yet to be determined. In such cases, typical training conditions should be used as the basis for making quality judgments.

Section II, Question 2b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section II, Question 3a. Frequency of Collective Work

How often has this task (or similar tasks) been performed by the collective?

1. Almost never.
2. Rarely.
3. Monthly.
4. Weekly.
5. Daily.

Definitions. Two collective tasks are considered *similar* if they share roughly the same processes and products/outcomes under differing conditions (e.g., applying the MDMP to different types of mission).

Note, to facilitate retention, collective work must develop cohesion and shared situation awareness. Therefore, for the purposes of this question, a task should be considered as performed frequently by the collective only if the *majority* of required personnel are present *each time*. This means that your response to this question should be affected by actual or typical (if actual is unknown) personnel availability, particularly in key leader positions.

Note, a collective task for which the majority of required personnel are present each time may be performed often or only under rare conditions, depending on the nature of the task.

Section II, Question 3b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section II, Question 4a. Quality/Use of SOP

How much have the SOP applicable to this task reduced the difficulty of coordination?

1. Existing SOP for this task don't get used/There are no SOP.
2. SOP *somewhat reduced* the difficulty of coordinating.
3. SOP *significantly reduced* the difficulty of coordinating.
4. SOP *totally* automated the coordination process.

Definitions: Some examples of SOP that reduce the difficulty of coordination include information management SOP and digital SOP.

Note units smaller than companies may have informal SOP to facilitate coordination. If so, they should be considered when answering this question.

Note also that SOP must be *used* to reduce coordination demands. Actual or typical (if actual is unknown) usage of SOP should be taken into consideration when answering this question.

Section II, Question 4b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

SECTION III – External Factors

The four questions in this section ask about factors outside of the unit's sphere of influence as they relate to conditions and technologies required to perform the task of interest (e.g., the usability of technology). In general, these questions ask: "***In your unit**, how do characteristics of the external environment influence the performance of the task in question?*" Therefore, external factors are to be judged using knowledge that relates to the particular unit to be trained and the task of interest.

Section III, Question 1a. Information Display User Friendliness

On average, how user friendly are the information displays that are most important to performing this task?

1. Difficult to use information displays actually make this task harder.
2. Somewhat usable.
3. Largely usable.
4. Completely user friendly/Not applicable (digital displays are not used at this echelon).

Definitions. *User friendly information displays* are easy to use, not overly complex, and place minimal requirements on users to train on, locate, or remember what interface features do and how they work together. They also do not require users to override their natural inclinations to use interface features correctly. Hard to use interfaces can actually make a task more difficult to perform without error. When judging information displays, consider the overall impact of user friendliness, specifically for those most important to executing the task.

Note that collective procedures may be developed and used as workarounds to for hard-to-use information displays. These should *not* be considered when answering this question.

Section III, Question 1b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section III, Question 2a. Available Technology

How frequently has the technology involved in performing this task changed?

1. Several times a year.
2. Once a year.
3. Every couple of years.
4. Never/Not applicable (No technology is used to perform this task).

Definitions. When answering this question it is important to consider how often the technology in the unit changes. For example, if the technology changes several times a year, but unit receives new technology once a year, then "once a year" should be selected.

Section III, Question 2b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section III, Question 3a. Technology Reliability

How reliable is the technology involved in performing this task?

1. Totally unreliable.
2. Somewhat unreliable.
3. Somewhat reliable.
4. Totally reliable/Not applicable.

Definitions. The reliability of technology is dependent on a number of factors, including:

- The degree to which the technology performs to government operational readiness specifications.
- The ruggedness of the technology in the conditions typical of field use.
- The degree to which the technology has been tested to ensure its full functionality with other software/equipment and for use in all operations.
- The availability of auxiliary or supporting equipment necessary to make the technology operate effectively.
- The degree to which the technology is practically useful to Soldiers.

Section III, Question 3b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Section III, Question 4a. Information Overload

What level of information overload does the leadership or overall collective performing this task typically operate under?

1. High levels of information overload.
2. Moderate levels of information overload.
3. Low levels of information overload.
4. No information overload.

Definitions. *Information overload* pertains to the sheer amount of information pushed to the leader or overall collective at the time of the task performance. Information overload reduces the ability to learn, perform, and retain most tasks, and may come from numerous sources, including:

- Multiple command and control information displays that must be used to perform missions.
- Frequent changes to technology and/or tactics.
- Highly complex missions with high-stakes outcomes.

Section III, Question 4b. Confidence Rating

How confident are you that your rating is accurate?

1. Totally confident.
2. Somewhat confident.
3. Not at all confident.

Appendix D

Individual and Collective Tasks Selected[3]

Individual Tasks – Unattended Ground Sensors (UGS)

Emplace tactical unattended ground sensors (T-UGS) – FCS007-632-6004
Emplace urban unattended ground sensor – FCS007-642-1002
Prepare a tactical unattended ground sensor (T-UGS) system for operation – FCS007-632-6002
Recover deployed tactical unattended ground sensor system – FCS007-632-6006
Recover deployed urban unattended ground sensor system – FCS007-642-1004
Prepare a tactical unattended ground sensor (T-UGS) plan – FCS007-632-6001
Prepare urban unattended ground sensor system for operation – FCS007-642-1001

Individual Tasks – Non-Line-Of-Sight Launch System (NLOS-LS)

Command and control NLOS-LS platoon/firing section movement operations
Direct and control NLOS-LS platoon/section operations
Change NLOS-LS operational modes – FCS007-511-0011
Conduct fire mission operations – FCS007-511-0016
Conduct NLOS-LS recovery procedures – FCS007-511-0018
Conduct preventative maintenance checks and services – FCS007-511-0020
Deliver embedded training – FCS007-511-0019
Install software – FCS007-511-0008
Load radio cryptographic and global positioning system keys – FCS007-511-0009
Maintain missile computer and communication system – FCS007-511-0024
Maintain power source – FCS007-511-0015
Navigate the interactive electronic technical manual – FCS007-511-0007
Perform basic removable control panel operations – FCS007-511-0003
Perform initialization procedures – FCS007-511-0004
Position container launch unit covers for operations – FCS007-511-0002
Prepare NLOS-LS for aerial transportation – FCS007-511-0013
Prepare NLOS-LS for a fire mission – FCS007-511-0014
Prepare NLOS-LS for ground transportation – FCS007-511-0012
Replenish NLOS-LS all up round – FCS007-511-0021
Set NLOS-LS system defaults – FCS007-511-0010

Collective Tasks – NLOS-LS

Perform reconnaissance operations NLOS-LS platoon
Conduct an air assault artillery raid
Conduct an artillery raid
Conduct air assault operations
Conduct an NLOS-LS fire mission

[3] Tasks listed in this appendix without task numbers had not been assigned task numbers at the time the study was conducted.

Coordinate container/launch unit (CLU) resupply
Conduct CLU reload operations
Perform CLU hangfire procedures
Conduct emergency missions
Establish firing capability at the firing position
Move a NLOS-LS platoon/firing section
Occupy a tactical assembly area (NLOS-LS)
Perform a survivability move – NLOS-LS Platoon/Section
Conduct occupation of position area NLOS-LS
Prepare the CLU for sling load operations
Transport a CLU
Prepare CLU for attended operations
Prepare CLU for unattended operations
Manage and submit NLOS-LS section reports

Collective Tasks – Assault

Staff Level

Implement the intelligence, surveillance, and reconnaissance plan – 17-1-1002.17-ACSQ
Conduct an air assault artillery raid

Company Level

Conduct consolidation/reorganization activities – 12-2-C021.17-D0KC
Conduct fire and movement – 71-2-0222-17-D0KC
Conduct reconnaissance handover – 71-2-4025.17-D0KC
Assault an enemy position – 71-2-0220-17-D0KC
Conduct troop leading procedures – 71-2-0065.17-D0KC
Conduct roadblock/checkpoint operations – 17-3-2324.17-D0KC
Cordon and search – 71-2-2027.17-D0KC
Conduct a screen – 71-2-0312.17-D0KC
Defend a battle position – 71-2-2603.17-D0KC
Clear a built-up area – 71-2-2025.17-D0KC
Conduct fire missions (fire support team)

*Platoon/Squad Leve*l

Assault an enemy position – 17-3-0220.17-KPLT
Rearm/Resupply – 17-3-1030.17-RECP
Conduct rearm/resupply operations – 17-3-0601.17-KPLT
Perform consolidation and reorganization – 17-3-2010.17-RECP
Conduct consolidation and reorganization – 07-3-5009.P
Search a building – 07-3-1414.P
Conduct deliberate occupation of a platoon battle position – 17-3-2602.17-KPLT
Conduct a platoon defense – 17-3-2605.17-KPLT

Conduct overwatch and/or support by fire – 07-3-1252.P
Conduct an attack – 07-3-1009.P
Conduct troop leading procedures – 17-3-0065.17-RECP
Conduct troop leading procedures – 07-3-5036
Establish an observation post – 17-3-1039.17-KPLT
Conduct target acquisition – 17.3-4017.17-RECP
Assault a building – 07-3-1000.P
Conduct a screen – 17-3-1023.17-RECP
Conduct a screen – 07-3-1144
Destroy an inferior force – 17-3-2450.17-KPLT
Conduct roadblock/checkpoint operations – 17-3-2324.17-RECP
Establish an observation post – Antiarmor/Infantry Reconnaissance Platoon/Squad
Conduct a defense – 07-3-1054
Conduct a reconnaissance handover – 17-3-4025.17-RECP

Appendix E

TLP Observer Checklist

Troop Leading Procedures Observer Checklist

The purpose of this checklist is to focus observer/controller attention as troop leading procedures (TLPs) are carried out. The foundation of this checklist is the assumption that GO/NO GO ratings (especially for collective tasks) paint a simpler picture of performance than may be reflected in other methods of performance assessment. In other words, some commanders and units may be "more GO" than others. This checklist is designed to capture the levels of skill development that reflect different "degrees of GO."

This checklist is based on scientific theory regarding what people can do at each of five different levels of skill development: (1) novice; (2) advanced beginner; (3) competent; (4) proficient; and (5) expert.

Going from novice to expert, each level of skill development reflects the refinements in behavior that occur as experience and knowledge are acquired. These refinements involve greater ability to sift relevant and irrelevant information, to perceive complex patterns, to prioritize information and tasks, and to direct behavior simultaneously towards short- and long-term goals.

This checklist is designed to support—with minimal modification—the observation of TLPs as conducted by a variety of unit types (e.g., armor, infantry reconnaissance, etc.), echelons, training environments, and training scenarios. Each of the high-level steps involved in TLPs is listed in the checklist along with:

(1) a general description of the step
(2) a general, but detailed description performance of the step associated with each of the 5 levels of expertise
(3) the aspects of the training scenario that, if known ahead of time, would help the observer to understand what each level of expertise "looks like" in a given exercise.

The intent is that this checklist will be used by military subject matter experts or by non-experts who have several years' experience with the military and observing military exercises. The checklist does not include specific exercise scenario details as part of its rating criteria, but observers should spend time familiarizing themselves with the exercise scenario they will be observing before using the checklist so that they will have the best understanding of what they should be looking for.

To use this checklist, first read through the TLP steps and their associated rating criteria. Review these together with the exercise scenario materials. While observing performance, use the following pages of the checklist to focus your observations during the different steps of TLPs. When you observe a particular behavior, mark the appropriate check box. The boxes you check will be combined to generate an overall TLP "score."

Maintain Situational Understanding

The platoon leader or company commander gains and/or maintains situational understanding using information that is gathered from FBCB2 (if applicable), FM communications, maps, intelligence summaries, SITREPs, and/or other available information sources. Situational understanding is used throughout troop leading procedures as the platoon leader or company commander conducts mission analysis, conducts reconnaissance, and refines his plan.

1 (Novice)	2	3	4	5 (Expert)
☐ Unit is unable to sift relevant from irrelevant information ☐ PL/CO CDR experiences information overload ☐ Heavy reliance is placed on FMs, handbooks, etc., to leverage C2 equipment and other sources of information ☐ Key sources of information that are not explicitly brought to the unit's attention (e.g., O/C prompts) are neglected ☐ Reporting among those involved in maintaining SU is either incomplete or excessive	☐ Unit is partially able to sift relevant from irrelevant information for certain aspects of the mission, based mostly on doctrinal understanding (not experience) ☐ Understanding how to leverage information resources is partially doctrine-based and partially experience-based ☐ Some key sources of common information are leveraged ☐ Reporting among those involved in maintaining SU is more targeted, but lacks analysis	☐ Unit sifts relevant from irrelevant information based on a pre-formed, short-range concept of the mission ☐ Leveraging SU resources reflects the immediate mission concept ☐ Obtaining SU involves the resources and personnel believed necessary to address the immediate mission concept ☐ Reporting among those involved in maintaining SU is targeted, but doesn't address possible information needs that arise outside of the immediate mission concept	☐ Unit sifts relevant from irrelevant information based on a long-term objective that includes the immediate circumstances and mission concept ☐ Leveraging SU resources reflects interest in both short- and long-term objectives ☐ Obtaining SU involves resources and personnel involved in the scope of the immediate and larger picture ☐ Capability to leverage C2 equipment to inform the long-term vision is well-developed ☐ Reporting goes to the right people and involves analysis (where applicable)	☐ Unit rapidly sifts relevant from irrelevant information based on a long-term or larger-picture mission concept that includes the immediate circumstances ☐ C2 equipment and information resources are maximally leveraged to inform SU over the short and long term ☐ Reporting among those involved in maintaining SU is rapid, targeted, and effective

Guidance for reviewing the exercise scenario: To understand what each of these behavior patterns "looks like," it will be helpful to know ahead of time: (1) What information sources are available in the exercise; (2) What information must be sought to address immediate objectives and the larger picture; (3) What information resources contain that information; and (4) Who needs to share information with whom in order to maintain situational understanding.

Issue a WARNO

The PL or CO CDR receives an OPORD or a FRAGO from higher and issues a WARNO to the platoon or company using FBCB2, FM radio, or other tactical means. The WARNO initiates the planning and/or preparation of subordinate units.

1 (Novice)	2	3	4	5 (Expert)
☐ PL or CO CDR includes too much or too little information in the WARNO and instructions are vague ☐ PL or CO CDR relies heavily on FM or other doctrinal resources to determine what should be included in the WARNO ☐ Method of sending WARNO does not take into account the best way to get the information to those who need it and is not timely	☐ PL or CO CDR includes key mission information in the WARNO, but instructions are somewhat vague ☐ PL or CO CDR relies on a combination of doctrine and experience to know what to include in the WARNO ☐ Method of sending the WARNO takes into account those who need it, but may miss some key people and is not timely	☐ PL or CO CDR includes key mission information in the WARNO and gives clear instructions to subordinate units ☐ PL or CO CDR builds a WARNO that is based on experience and relates to the characteristics of the situation ☐ Method of sending the WARNO is context-sensitive, but may not be timely	☐ PL or CO CDR develops a complete, targeted WARNO ☐ Method of sending the WARNO is targeted, but is not automatic	☐ PL or CO CDR rapidly develops a complete, targeted WARNO on the basis of experience

Guidance for reviewing the exercise scenario: To understand what each of these behavior patterns "looks like," it will be helpful to know ahead of time: (1) what context the WARNO should be sensitive to; (2) what the timeframe for issuing WARNOs should be; and (3) who needs to receive the WARNO.

Mission Analysis

The PL or CO CDR conducts a timely and effective mission analysis using the factors of METT-TC to guide observation and information organization.

1 (Novice)	2	3	4	5 (Expert)
☐ PL or CO CDR neglects to consider some METT-TC factors ☐ Unit experiences information overload ☐ PL or CO CDR does not consider elements of his analysis as integrated, dependent factors, but as isolated variables that all are of equal importance to the mission ☐ PL or CO CDR does not realize when he doesn't understand the 2-up intent ☐ PL or CO CDR relies heavily on FMs, handbooks, etc., to determine what his specified and implied tasks are ☐ Coordination with others, where applicable, is not prioritized or timely	☐ PL or CO CDR considers each of the METT-TC factors, but misses some key information ☐ Some information overload is experienced when there is difficulty prioritizing, making the analysis untimely ☐ PL largely treats information as isolated variables, rather than dependent components of a complex pattern ☐ PL or CO CDR relies on a combination of FMs, handbooks, etc., and experience to conduct the mission analysis ☐ PL or CO CDR recognizes when he doesn't understand the 2-up intent, but doesn't necessarily do something to clarify misunderstanding; (may go to sources other than the key source to clarify understanding) ☐ Coordination with others, where applicable, is not well prioritized or timely	☐ PL or CO CDR considers the key information relating to each of the METT-TC factors based on the pre-formed mission concept ☐ Some complex patterns in the analysis are recognized ☐ New information that may suggest modifications to the mission concept is overlooked or not used to form SU of the larger picture ☐ Coordination with others, where applicable, reflects adherence to the mission concept, even when circumstances call for adjustment	☐ PL or CO CDR considers the key information relating to each of the METT-TC factors based on a short- and longer-term mission concept, but not automatically ☐ Complex patterns in the analysis are recognized, but may have time costs ☐ PL or CO CDR modifies mission concept to address new information ☐ PL or CO CDR generates clear implied and specified tasks, but may not be timely ☐ Coordination with others, where applicable, is prioritized, but may not be fully timely	☐ PL or CO CDR rapidly considers the key information relating to each of the METT-TC factors and automatically perceives complex patterns based on a short- and longer-term mission concept ☐ PL or CO CDR modifies mission concept to address new information ☐ PL or CO CDR rapidly generates clearly specified and implied tasks ☐ Coordination with others, when applicable, is prioritized and timely

Guidance for reviewing the exercise scenario: To understand what each of these behavior patterns "looks like," it will be helpful to know ahead of time: (1) what METT-TC information should be considered, what complex patterns are present, and how information should be prioritized; (2) what clearly stated implied and specified tasks for the mission should be; and (3) what coordination with others to conduct the mission is required.

Make a Tentative Plan

PL or CO CDR makes a tentative plan based on a detailed mission analysis (see above) and situational understanding (see above). The tentative planning process includes COA development, COA analysis, COA comparison, and a tactical decision.

1 (Novice)	2	3	4	5 (Expert)
☐ PL or CO CDR develops a COA(s) that reflects relatively little understanding of the METT-TC factors and their interaction ☐ Little time is spent on mission analysis relative to COA development ☐ COA analysis does not take into account the likely and most dangerous events that could occur during the mission ☐ Where events are considered, only a surface level consideration is applied ☐ COA analysis does not synchronize the mission or produce decision points ☐ Criteria used to compare COAs (if developed, and where applicable), are vague and do not reflect issues of importance to the mission ☐ FMs, handbooks, etc., are relied upon heavily to conduct the planning process; Process is labor intensive and untimely	☐ PL or CO CDR develops a COA(s) that reflects some understanding of the METT-TC factors, but not their complex interdependency ☐ COA analysis takes into account the most likely and most dangerous events that could occur during the mission ☐ Largely surface level consideration is applied to events during COA analysis, but for common events or events within the PL's or CO CDR's limited experience range, more detail is considered ☐ COA analysis doesn't synchronize the mission or produce decision points ☐ Criteria used to compare COAs (if developed, and where applicable), are vague ☐ A combination of experience and doctrine is used to conduct the planning process ☐ Process is labor intensive and untimely	☐ PL or CO CDR develops a COA(s) that reflects good understanding of the METT-TC factors as well as some understanding of their complex interdependence ☐ COA analysis considers key events at a more developed level of detail, although emphasis is placed on confirming the existing pre-conceived mission concept rather than exploring all options ☐ COA analysis synchronizes the mission and produces decision points, but outcome may not lead the PL or CO CDR to alter his existing mission concept ☐ Criteria used to compare COAs (where applicable) are fairly clear, but may not adequately "test" the COAs ☐ Planning process is experientially based ☐ Process is not rapid, but more timely than for novices and advanced beginners	☐ PL or CO CDR develops a COA(s) that reflects the interdependencies among the METT-TC factors ☐ COA analysis leads to a synchronized plan with decision points identified ☐ COA comparison (where applicable) adequately "tests" each COA ☐ COA analysis and comparison leads to modification of the initial mission concept ☐ Planning process is experientially based, but not necessarily rapid	☐ PL or CO CDR rapidly develops a COA(s) that reflects the complex interdependencies among the METT-TC factors ☐ More time is spent during mission analysis than COA development ☐ COA analysis leads to a well-synchronized plan with decision points identified ☐ COA comparison (where applicable) is rapid and adequately "tests" each COA ☐ COA analysis and comparison leads to modification of the initial mission concept ☐ Planning process is experientially based, rapid and timely

Guidance for reviewing the exercise scenario: To understand what each of these behavior patterns "looks like," it will be helpful to know ahead of time: (1) what a well-developed COA (or set of COAs) for the mission would look like; (2) what a refined plan for the mission based on COA analysis would look like; and (3) what effective COA comparison criteria for the mission would be.

Initiate Movement

PL or CO CDR initiates movement IAW the WARNO, OPORD, FRAGO, and/or unit SOP. (There may be a need to initiate movement immediately following the issuance of the WARNO).

1 (Novice)	2	3	4	5 (Expert)
☐ Movement tasks are not prioritized, all having seemingly equal importance to the PL/CO CDR ☐ Security measures, control measures, and tactical decisions reflect surface-level understanding of the movement subtask ☐ Doctrinal or other templates are used to frame decision making and action, rather than experience and sensitivity to the context of the operational environment and mission	☐ Movement tasks are prioritized according to a generalized doctrinal solution, and not contextualized to the environment or mission ☐ Security measures, control measures, and tactical decisions reflect slightly more than surface-level understanding of the movement subtask ☐ Movement initiation and coordination is partially experience-based and partially based on doctrinal or other templates, demonstrating partial sensitivity to the operational environment and mission	☐ Movement tasks are prioritized, but are not sensitive to changes in the environment or information available ☐ Security measures, control measures, and tactical decisions reflect detailed understanding of the plan, but where the environment differs from expected, they are not adjusted ☐ Movement initiation and coordination is experience-based, but is resistant to rapid adaptation	☐ Movement tasks are prioritized according to the factors of mission, enemy, and terrain and are changed as the situation dictates, if not in a rapid manner ☐ Security measures, control measures, and tactical decisions reflect detailed understanding of the plan and of the operational environment and can be adjusted, albeit not rapidly ☐ Movement initiation and coordination is experience-based and may be adapted, just not rapidly	☐ Prioritized movement tasks are executed rapidly and are swiftly adapted if the situation changes ☐ Security measures, control measures, and tactical decisions are rapidly executed and flexible ☐ Movement initiation and coordination is experience-based and automatic

Guidance for reviewing the exercise scenario: To understand what each of these behavior patterns "looks like," it will be helpful to know ahead of time: (1) what the initial movement requirements (and associated orders, instructions, etc.) for the scenario should be; (2) what (if any) "hooks" or "trigger events" are built into the scenario to require rapid tactical decision making during initial movement and what the appropriate decisions would be; and (3) what the prioritization of movement tasks should be.

Conduct Reconnaissance

PL or CO CDR conducts reconnaissance as time and resources allow. Reconnaissance may be map-based or personal and may require coordination with other units.

1 (Novice)	2	3	4	5 (Expert)
☐ PL's or CO CDR's selection, prioritization, and use of recon techniques and assets reflects lack of understanding of the resources and time available for recon and the remaining information requirements for the mission ☐ All recon targets are considered equally important and do not address the remaining critical information requirements ☐ PL or CO CDR fails to recognize when situation calls for a request for additional recon assets ☐ Decisions are not experience-based, but follow some generalized template of how a recon should be conducted	☐ PL's or CO CDR's selection, prioritization, and use of recon techniques and assets reflects surface-level understanding of the resources and time available for recon and the remaining information requirements for the mission ☐ Commonly identified recon targets are given top priority but these targets may not necessarily address the remaining critical information requirements ☐ Decisions are partially experience-based, but not necessarily context-sensitive	☐ PL's or CO CDR's selection, prioritization, and use of recon techniques and assets reflects understanding of the resources and time available and knowledge of the remaining information requirements for the mission ☐ Recon targets are prioritized to address the remaining critical information requirements ☐ PL or CO CDR recognizes when situation calls for a request for additional recon assets ☐ Decisions are experience-based, but not necessarily context-sensitive	☐ PL or CO CDR selects a targeted, effective recon technique, although this selection may not be rapid ☐ Recon priorities address the remaining critical information requirements in the order of importance to the mission ☐ PL or CO CDR recognizes when situation calls for a request for additional recon assets ☐ Decisions are experience-based and context-sensitive, but not necessarily timely	☐ PL or CO CDR rapidly selects a timely and effective recon technique ☐ Recon priorities address the remaining critical information requirements in the order of importance to the mission ☐ PL or CO CDR immediately recognizes when the situation calls for a request for additional assets ☐ Decisions are experience-based, context-sensitive and automatic

Guidance for reviewing the exercise scenario: To understand what each of these behavior patterns "looks like," it will be helpful to know ahead of time: (1) what the expected recon requirements (method and critical information requirements) for the scenario are; and (2) what resources are available to conduct recon in the exercise.

Complete the Plan

PL or CO CDR completes the plan, adjusting it based on the results of the reconnaissance and refining it to meet the mission requirements and commander's intent.

1 (Novice)	2	3	4	5 (Expert)
☐ PL or CO CDR doesn't recognize when information from the recon requires an adjustment to the plan ☐ PL or CO CDR doesn't formulate a plan that complies with the commander's intent and/or doesn't recognize when the plan doesn't fully meet the mission requirements and commander's intent	☐ PL or CO CDR occasionally recognizes when information from the recon requires an adjustment to the plan, but doesn't necessarily know how to adjust the plan ☐ PL or CO CDR does not intentionally formulate a plan that complies with mission requirements and commander's intent (it's a happy accident) ☐ PL or CO CDR occasionally recognizes when the plan doesn't fully meet the mission requirements and commander's intent, but doesn't know how to adjust it	☐ PL or CO CDR recognizes when information from the recon requires an adjustment to the plan, but doesn't make the adjustments or only makes surface-level adjustments ☐ PL or CO CDR generally formulates a plan that complies with mission requirements and commander's intent ☐ PL or CO CDR recognizes when the plan doesn't fully meet the mission requirements and commander's intent, but doesn't adjust it or only makes surface-level adjustments	☐ PL or CO CDR recognizes when information from the recon requires an adjustment to the plan, and makes the adjustments ☐ PL or CO CDR formulates a plan that complies with the mission requirements and commander's intent, but it's necessarily not a timely process ☐ Making adjustments isn't necessarily timely or always effective	☐ PL or CO CDR automatically recognizes when information from the recon requires an adjustment to the plan, and rapidly makes the adjustments ☐ PL or CO CDR rapidly formulates a plan that complies with the mission requirements and commander's intent ☐ Making adjustments is timely and effective

Guidance for reviewing the exercise scenario: To understand what each of these behavior patterns "looks like," it will be helpful to know ahead of time: (1) what a fully adjusted and effective plan, given the exercise scenario, would be.

Issue Orders and Instructions

PL or CO CDR issues orders and instructions to include ROE and ROI.

1 (Novice)	2	3	4	5 (Expert)
☐ PL or CO CDR doesn't recognize when subordinates don't understand the mission, his commander's intent, concept of operation and/or assigned tasks ☐ Orders and instructions do not take into account what subordinates need to know ☐ PL or CO CDR does not understand how ROE/ROI relate to how the mission should be conducted	☐ PL or CO CDR sometimes recognizes when subordinates don't understand the mission, commander's intent, concept of operation and/or assigned tasks, but doesn't know what to do about it ☐ Orders and instructions partially take into account what subordinates need to know ☐ PL or CO CDR has only a surface-level understanding of how ROE/ROI relate to how the mission should be conducted	☐ PL or CO CDR recognizes when subordinates don't understand the mission, commander's intent, concept of operation and/or assigned tasks, but doesn't address the shortfall ☐ Orders and instructions generally take into account what subordinates need to know ☐ PL or CO CDR has a working understanding of how ROE/ROI relate to the mission, but only the obvious connections are discussed	☐ PL or CO CDR recognizes and assists when subordinates don't understand the mission, commander's intent, concept of operation and/or assigned tasks ☐ Orders and instructions take into account what subordinates need to know, but are not rapidly generated ☐ PL or CO CDR has an in-depth understanding of how ROE/ROI relate to the mission, and articulates both obvious and subtle connections	☐ PL or CO CDR automatically recognizes and assists when subordinates don't understand the mission, commander's intent, concept of operation and/or assigned tasks ☐ Orders and instructions take into account what subordinates need to know and are rapidly generated ☐ PL or CO CDR has an in-depth understanding of how ROE/ROI relate to the mission, and automatically articulates both obvious and subtle connections

Guidance for reviewing the exercise scenario: To understand what each of these behavior patterns "looks like," it will be helpful to know ahead of time: (1) what the ROE/ROI are and how they relate to the mission; (2) what signs indicate to a leader that the mission, commander's intent, concept of operation, and/or assigned tasks are not fully understood; and (3) what subordinates need to know in the orders and instructions, given the scenario.

Supervise Preparations and Refine Order

PL or CO CDR conducts a rehearsal, if possible, and conducts inspections.

1 (Novice)	2	3	4	5 (Expert)
☐ Rehearsal addresses all aspects of the mission equally instead of focusing on those aspects that are most difficult or most important ☐ Inspections are conducted based on a doctrinal or some other form of general checklist, rather than being specific to the mission requirements and context	☐ Rehearsal reflects some prioritization of the mission tasks, but emphasis is still placed on some relatively simple or relatively unimportant tasks ☐ Inspections are partially based on a doctrinal or some other form of general template, and are partially specific to the mission requirements and context ☐ Rehearsal and inspections are disorganized and untimely	☐ Rehearsal prioritization reflects some, but not all of the most challenging and most important mission tasks ☐ Inspections are specific to the surface-level mission requirements and context ☐ PL or CO CDR does not use information gathered during rehearsal or inspections to refine the order or instructions ☐ Rehearsal and inspections are somewhat disorganized and untimely	☐ Rehearsal prioritization reflects the most challenging and most important mission tasks, but the prioritization process isn't rapid ☐ Inspections are specific to the mission requirements and context, but are not rapid ☐ PL or CO CDR uses information gathered during rehearsal or inspections to refine the order, but refinements are not necessarily made quickly, so some may be lost due to time pressure ☐ Rehearsal and inspections are organized but not necessarily rapid	☐ Rehearsal prioritization is rapid, and reflects the most challenging and most important mission tasks ☐ Inspections are specific to the mission requirements and context, and conducted rapidly ☐ PL or CO CDR uses information gathered during rehearsal or inspections to rapidly refine the order ☐ Rehearsal and inspections are organized and rapid

Guidance for reviewing the exercise scenario: To understand what each of these behavior patterns "looks like," it will be helpful to know ahead of time: (1) what the most challenging and most important mission tasks are; and (2) what the priorities for inspection should be given the exercise scenario and mission requirements.